THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXI February, 1940 Number 2

THE VALUE OF GROUP TEST IQ'S FOR PREDICTION OF PROGRESS BEYOND HIGH SCHOOL

NOEL KEYS

University of California

In the great majority of schools, individual Binet tests remain a luxury reserved for the occasional problem child. Group tests, however, are given most pupils at some stage in their elementary-school careers, and in scientifically administered systems such records are available for practically the entire student body by junior high school. The value of group test IQ's for counselling with a view to advanced schooling should, therefore, be at a maximum around the eighth or ninth grade. Counsellors, however, lack the necessary complementary information as to the intelligence needed for success in various types of schooling beyond high school. Nearly a quarter of a century since the introduction of group tests our knowledge even of the IQ normal for college and university work is still meager in the extreme. For this, two circumstances are responsible. The literature on prediction of college success from intelligence measurements is extensive,[6] but confined almost entirely to tests administered at or near time of admission to college. Aptitude tests designed for high-school seniors seldom yield results in terms of intelligence quotients. On the other hand, long term follow-ups of the later careers of pupils in relation to IQ's actually obtained in elementary or junior high school grades are rare.

PREVIOUS STUDIES OF LONG-RANGE PREDICTION

Such long-range studies as the literature affords cast grave doubt, not only on the value of such IQ's for purposes of prediction, but also on previous estimates of the intelligence requisite for university success. Adams[1] found that of fifteen hundred five pupils in the white elementary schools of a Texas university town in 1926, those who later

entered college had, when in the fourth to the sixth grade, averaged around the upper quartile of their classes in IQ on the National Intelligence Test. Other figures in his report indicate the median IQ of this group to have been approximately 112, as compared with 106 or 107 for all those who graduated from high school. For the three hundred eighty-seven known to have attended college, the correlation of these early IQ's with scholarship average in freshman year was only .38 ± .03.

Portenier[5] followed thirty pupils constituting the bright and accelerated section and the twenty-two making up a dull and retarded section in the ninth grade of Lincoln (Nebraska) High School in 1923–1924. The former had group test IQ's ranging from 122 to 146 with a median of 133, while the latter ranged from 67 to 108, with a median of 93. The results showed:

	Per cent of bright	Per cent of dull
Completing four years in high school	90	68
Completing one year in university	87	50
Completing four years in university	63	32
Graduating from university with honors	23	0

Byrns and Henmon[3] traced pupils given the National Intelligence Test in Grades III to VIII of the Madison (Wisconsin) schools. For the one hundred fifty who eventually entered the University of Wisconsin in that city, these elementary school IQ's were found to have predicted first-semester scholarship to the extent of $r = .454$. This matches the correlation of .459 between percentile status on the Ohio Psychological Examination given in senior year of high school and the college marks, and slightly exceeds the median correlation of .44 reported by Segel[6] from many studies of the relationship between group test scores on college entrance and later academic success. The mean IQ of the university entrants, however, proved to have been only 109.3 ± 0.36. This is even lower than Adams' findings in the Texas situation, and contrasts strangely with Ellis's[4] conjecture in 1928 that for a Stanford-Binet IQ of 110, the tenth grade is "the final limit to which education may profitably be carried with our present curriculum," and that 140 is needed for university work. Counsellors may well ask which figure should be trusted.

PROBLEM AND PROCEDURE

The present study represents an attempt to ascertain the significance of group test IQ's obtained in junior-high-school years for the prediction of academic experience beyond high school. More particularly, it seeks to throw fresh light upon (*a*) the distribution of IQ's for those who proceed from high school to various types of institution, (*b*) the correlation between these IQ's and scholarship in the two institutions found to have received the greatest number, and (*c*) the probability of particular sorts of advanced schooling and of academic success therein for pupils of given IQ levels.

With the coöperation of Principal L. P. Farris, the records of all pupils leaving the Oakland (California) High School from midwinter 1928–1929 to midwinter 1933–1934 were examined. After eliminating the few transferring to other high schools, all those were selected for whom IQ's had been obtained on the Terman Group Test of Mental Ability at some time within Grades VII to X and before their fifteenth birthdays. The Terman Test was selected as the one most widely used at entering high school level, while the restriction as to age helped to avoid questions as to the validity of age norms in the later teens and the proper divisor to be used in computing quotients at these ages. The resulting group numbered eleven hundred twelve, and constituted a presumably random sampling of the entire student body. The direction of each pupil's advanced schooling, if any, was inferred from the institution or institutions to which the transcript of his record was sent. During the year 1938—from five to ten years after each had left high school—the scholastic success of those who had entered either the University of California or the Merritt Business School (maintained as a part of the Oakland city schools) was determined by a search of the records of these institutions. These two institutions accounted for three-fourths of the transcripts sent. No adequate attempt was made to verify the later academic success of the remaining one-quarter of transfers whose transcripts were scattered among some fifty other institutions. However, the records at Merritt and the University showed that in each case nearly ninety per cent of those sending transcripts had actually enrolled at some time. Moreover, the distributions of IQ's for those who sent transcripts but did not enter and for those who did enter showed almost identical means and sigmas (*cf.* Table II). Under these circumstances, it seemed justifiable to regard the IQ's of the boys and girls forwarding transcripts to a given

class of institution as fairly representative of those who proceed from Oakland High School to schools of that sort.

DISTRIBUTION OF ALUMNI AMONG VARIOUS TYPES OF SCHOOL

As shown in Table I, thirty-five per cent of the student body sent no transcripts of record. Since virtually all institutions call for this information, it is reasonable to suppose that for fully one-third of even

TABLE I.—PERCENTAGE OF HIGH-SCHOOL STUDENTS SENDING TRANSCRIPTS TO SCHOOLS OF VARIOUS TYPES

Destination of transcripts	Number of pupils		Percentage of pupils	
No transcript furnished		390		35.1
Non-degree-granting schools.				
Merritt Business School	214		19.2	
Other schools of business, nursing, or applied arts	48		4.3	
Junior colleges	51		4.6	
		313		28.1
Degree-granting institutions.				
University of California.				
Applicants who did not enter	35		3.1	
Students actually entering	290		26.1	
	325		29.2	
California state colleges	24		2.2	
Other colleges and universities	60		5.4	
		409		36.8
Total leaving Oakland High School		1,112		100.0

the somewhat select population studied, formal schooling ceased with high school. The few who may have secured admission to some school without a transcript would almost certainly be offset by the number who sent transcripts elsewhere without entering. Thirty-seven per cent of the total had sooner or later forwarded transcripts to degree-granting colleges or universities, and the remaining twenty-eight per cent to schools of business, trades, fine or applied arts, nursing, junior colleges, and other non-degree institutions.

DISTRIBUTION OF INTELLIGENCE QUOTIENTS ACCORDING TO TYPE OF SCHOOL

An examination of Table II discloses that:

(1) The mean IQ for the entire group studied was 107.9. The median was 107.7. This is somewhat higher than that of most other

TABLE II.—AVERAGE AND RANGE OF INTELLIGENCE QUOTIENTS OF STUDENTS TRANSFERRING TO SCHOOLS OF VARIOUS TYPES

	Number	Mean PE	Sigma	Range
Sending no transcripts	390	103.3 ± 0.43	12.7	70 to 138
Sending transcripts to non-degree institutions.				
Merritt Business School	214	104.3 ± 0.5	11.8	73 to 138
Other vocational schools	48	104.5 ± 1.1	11.8	81 to 132
Junior colleges	51	106.0 ± 1.2	13.2	82 to 136
Total	313	104.6 ± 0.46	12.0	73 to 138
Sending transcripts to degree-granting institutions but not entering University of California.				
California state colleges	24	108.0 ± 1.6	11.6	88 to 137
Other colleges and universities	60	113.2 ± 1.1	12.7	91 to 142
Transcripts to University of California but never registered	35	115.6 ± 1.3	11.7	92 to 143
Total	119	112.8 ± 0.77	12.5	88 to 143
Entering University of California.				
But failing to obtain the Junior Certificate	92	111.1 ± 0.7	10.5	86 to 144
Obtaining Junior Certificate but not bachelor's degree	44	115.2 ± 1.0	9.4	95 to 137
Receiving bachelor's degree				
Without honors	135	117.3 ± 0.6	10.1	86 to 140
With honors	19	125.2 ± 2.1	13.8	99 to 152
Total graduating	154	118.3 ± 0.6	11.0	86 to 152
Total University of California students	290	115.5 ± 0.44	11.1	86 to 152
Total leaving high school	1,112	107.9 ± 0.26	13.2	70 to 152

high schools in Oakland, and also higher than the median of 104 found in four California high schools studied in 1933–1934 by Bell and Proctor ([2], p. 692). It is, however, in keeping with the better than average social-economic character of the residential section in which the Oakland High School is located. This circumstance should not seriously affect the norms obtained in the present study for various types of advanced schooling, since Adams has shown that there is little relation between the median IQ in a given school and the IQ or scholarship of the particular pupils from that school who later enter university ([1], p. 63ff.).

(2) When the eleven hundred twelve pupils are grouped according to later academic destination, the mean IQ's form a gradually ascending series from 103.3 for those taking no transcripts up to 125.2 for those graduating with honors from the University of California.

(3) The average IQ of the two hundred ninety students actually entering the University proved to be 115.5 ± 0.44. This is more than three points above the Texas medians and six points higher than at Wisconsin. The differences, moreover, are highly significant statistically. The explanation might lie in (*a*) discrepancies in the IQ's yielded by the Terman as compared with the National Intelligence Test, (*b*) differing admission requirements of the three universities, or (*c*) the fact that the present quotients were obtained at junior high instead of elementary school level. At all events, 115.5 is much more nearly in accordance with previous estimates of the IQ typical of university undergraduates.

(4) The mean IQ of the one hundred fifty-four who succeeded in graduating from the University is 118.3, as compared with 112.4 for the one hundred thirty-six who entered but failed to graduate. This difference, also, is statistically reliable.

(5) Oakland High School students proceeding to non-degree-granting institutions average significantly lower than the population from which they are drawn. It is a little surprising to find that pupils transferring to business schools of various sorts have a mean IQ three and six-tenths points lower than the high school as a whole. This difference, though small, amounts to six times its probable error.

(6) Students proceeding to degree-granting institutions average well above the high-school student body. The differences are sufficient for practical certainty in the case of all groups excepting the twenty-four transfers to California state colleges. The last-named show a mean IQ of 108.0, practically identical with the total high-school group.

The low scores found in the case of transfers to junior colleges and state colleges is doubtless largely a consequence of the particular geographical location. The largest state university in America is open, free of tuition, only eight miles from Oakland High School. None of the state colleges or junior colleges is so conveniently accessible, and private colleges in the locality charge substantial fees. It is almost inevitable that ambitious students who can qualify for the University should tend to seek admission there, leaving to lesser institutions an unfavorable selection of those whose scholarship is for the most part inadequate for the University's requirements. In a district close to a state college or junior college and remote from a university, the showing might well be different. From a high school so situated the number of students of high IQ who enter the college or junior college would almost certainly be greater.

(7) Finally, the range of IQ's in each group is large and the overlapping great. Little can be said save that no student seeking admission to a degree-granting institution showed an IQ lower than 86, and none of those transferring to non-degree schools one higher than 138.

PREDICTION OF COLLEGE SCHOLARSHIP POSSIBLE FROM JUNIOR HIGH-SCHOOL IQ

Correlations between Terman IQ's and scholarship in university or business college were as follows:

Correlation of IQ with	Number	r	PE
Average of instructors' ratings in three or more courses completed at Merritt Business School	156	.15	± .054
Grade-point average in all undergraduate courses taken at University of California	279	.35	± .035

The ratings at Merritt Business School are not directly comparable with ordinary academic grades, representing rather the instructor's judgment as to a student's readiness for employment at the kind of work involved. This fact, together with the non-academic character of the courses, may account for the insignificant correlation with intelligence quotients. Despite this low observed relationship, however, the Merritt School has recently ruled that applicants with IQ below 90 shall be denied admission as bad risks except where high scholarship record or other objective evidence appears to warrant special action.

high schools in Oakland, and also higher than the median of 104 found in four California high schools studied in 1933–1934 by Bell and Proctor ([2], p. 692). It is, however, in keeping with the better than average social-economic character of the residential section in which the Oakland High School is located. This circumstance should not seriously affect the norms obtained in the present study for various types of advanced schooling, since Adams has shown that there is little relation between the median IQ in a given school and the IQ or scholarship of the particular pupils from that school who later enter university ([1], p. 63ff.).

(2) When the eleven hundred twelve pupils are grouped according to later academic destination, the mean IQ's form a gradually ascending series from 103.3 for those taking no transcripts up to 125.2 for those graduating with honors from the University of California.

(3) The average IQ of the two hundred ninety students actually entering the University proved to be 115.5 ± 0.44. This is more than three points above the Texas medians and six points higher than at Wisconsin. The differences, moreover, are highly significant statistically. The explanation might lie in (*a*) discrepancies in the IQ's yielded by the Terman as compared with the National Intelligence Test, (*b*) differing admission requirements of the three universities, or (*c*) the fact that the present quotients were obtained at junior high instead of elementary school level. At all events, 115.5 is much more nearly in accordance with previous estimates of the IQ typical of university undergraduates.

(4) The mean IQ of the one hundred fifty-four who succeeded in graduating from the University is 118.3, as compared with 112.4 for the one hundred thirty-six who entered but failed to graduate. This difference, also, is statistically reliable.

(5) Oakland High School students proceeding to non-degree-granting institutions average significantly lower than the population from which they are drawn. It is a little surprising to find that pupils transferring to business schools of various sorts have a mean IQ three and six-tenths points lower than the high school as a whole. This difference, though small, amounts to six times its probable error.

(6) Students proceeding to degree-granting institutions average well above the high-school student body. The differences are sufficient for practical certainty in the case of all groups excepting the twenty-four transfers to California state colleges. The last-named show a mean IQ of 108.0, practically identical with the total high-school group.

The low scores found in the case of transfers to junior colleges and state colleges is doubtless largely a consequence of the particular geographical location. The largest state university in America is open, free of tuition, only eight miles from Oakland High School. None of the state colleges or junior colleges is so conveniently accessible, and private colleges in the locality charge substantial fees. It is almost inevitable that ambitious students who can qualify for the University should tend to seek admission there, leaving to lesser institutions an unfavorable selection of those whose scholarship is for the most part inadequate for the University's requirements. In a district close to a state college or junior college and remote from a university, the showing might well be different. From a high school so situated the number of students of high IQ who enter the college or junior college would almost certainly be greater.

(7) Finally, the range of IQ's in each group is large and the overlapping great. Little can be said save that no student seeking admission to a degree-granting institution showed an IQ lower than 86, and none of those transferring to non-degree schools one higher than 138.

PREDICTION OF COLLEGE SCHOLARSHIP POSSIBLE FROM JUNIOR HIGH-SCHOOL IQ

Correlations between Terman IQ's and scholarship in university or business college were as follows:

Correlation of IQ with	Number	*r* PE
Average of instructors' ratings in three or more courses completed at Merritt Business School	156	.15 ± .054
Grade-point average in all undergraduate courses taken at University of California	279	.35 ± .035

The ratings at Merritt Business School are not directly comparable with ordinary academic grades, representing rather the instructor's judgment as to a student's readiness for employment at the kind of work involved. This fact, together with the non-academic character of the courses, may account for the insignificant correlation with intelligence quotients. Despite this low observed relationship, however, the Merritt School has recently ruled that applicants with IQ below 90 shall be denied admission as bad risks except where high scholarship record or other objective evidence appears to warrant special action.

Even with grade-point averages in university work, the correlation of .35 falls far short of the .45 obtained by Byrns and Henmon with

TABLE III.—DISTRIBUTION OF INTELLIGENCE QUOTIENTS AMONG STUDENTS TRANSFERRING TO SCHOOLS OF VARIOUS TYPES

	Intelligence quotients					
	70–84	85–94	95–104	105–119	120–139	140–151
Sending no transcripts	27	67	114	139	43	
Sending transcripts to non-degree institutions.						
Merritt Business School	11	28	78	75	22	
Other vocational schools	1	11	11	20	5	
Junior colleges	1	11	12	18	9	
Total	13	50	101	113	36	0
Sending transcripts to degree-granting institutions but not entering University of California.						
California state colleges	..	2	10	9	3	
Other colleges or universities	..	4	14	27	14	1
Transcripts to University of Calif., but never registered there	..	1	2	20	10	2
Total	0	7	26	56	27	3
Entering University of California.						
Failing to obtain the Junior Certificate	..	4	22	46	19	1*
Receiving Junior Certificate but not bachelor's degree	..	....	4	28	12	
Receiving bachelor's degree.						
Without honors	..	2*	12	63	57	1*
With honors	..	....	1*	6	8	4
Total University of California students	0	6	39	143	96	6
Total leaving Oakland High School	40	132	283	459	203	9

*See text for explanation of these exceptional cases.

first semester grades at the University of Wisconsin. It is, however, only slightly below the .37 found by the same investigators with four-

year average grades, and the .38 obtained by Adams in Texas. The general significance of a correlation of this order may be judged from an examination of Table III. From the data there summarized, it is possible to say, for example, that:

(1) Of the one hundred fifty-four students who graduated from the University of California, only two had IQ's below 95. Each of these is known to have come from a home in which a language other than English was regularly spoken, and that may have operated to depress unduly the IQ recorded in junior high school.

(2) Of the ninety-two students failing to qualify for the Junior Certificate covering the first two years of university work, only one had an IQ above 140. This boy, though brilliant, suffered from character defects which led to his dismissal for a combination of disciplinary reasons and scholarship deficiency after several years in the University.

(3) Of the one hundred thirty-five students graduating from the University of California without honors, only one had an IQ as high as 140, and that one, a girl, fell short by only 0.02 of the grade-point average of 2.00 required for honors.

(4) Of the nineteen students graduating from the University of California with honors, only one had a recorded IQ below 106. This was a girl who registered an IQ of 99 in junior high school when she had been only eighteen months in America. Her scholarship throughout high school showed an almost straight *A* record, sufficient to demonstrate that the intelligence rating was invalid in her case. In the University she received her B.A. with honors in modern languages.

THE ACADEMIC PROSPECTS OF PUPILS OF VARIOUS IQ LEVELS

Figure 1 compares the percentage of persons of various IQ groups who proceeded to particular academic destinations. In Table IV the same data are presented in terms of probability or chances. By running the eye down the appropriate column, it is possible to read the chances in one hundred that a junior-high-school pupil of the IQ level shown will attain to any particular state of schooling, under the conditions of the present study.[1] For example:

[1] Note that the probabilities as indicated apply to children of about the ninth-grade level, rather than to college entrants or the like. The chances that a university freshman of 90 or 100 IQ, for example, will succeed in graduating should be much better than the chances for the average junior-high-school pupil of that mentality. By the very fact of having entered the university the former has demonstrated the possession of qualities or circumstances which place him among a very small and select minority of his equals in intelligence.

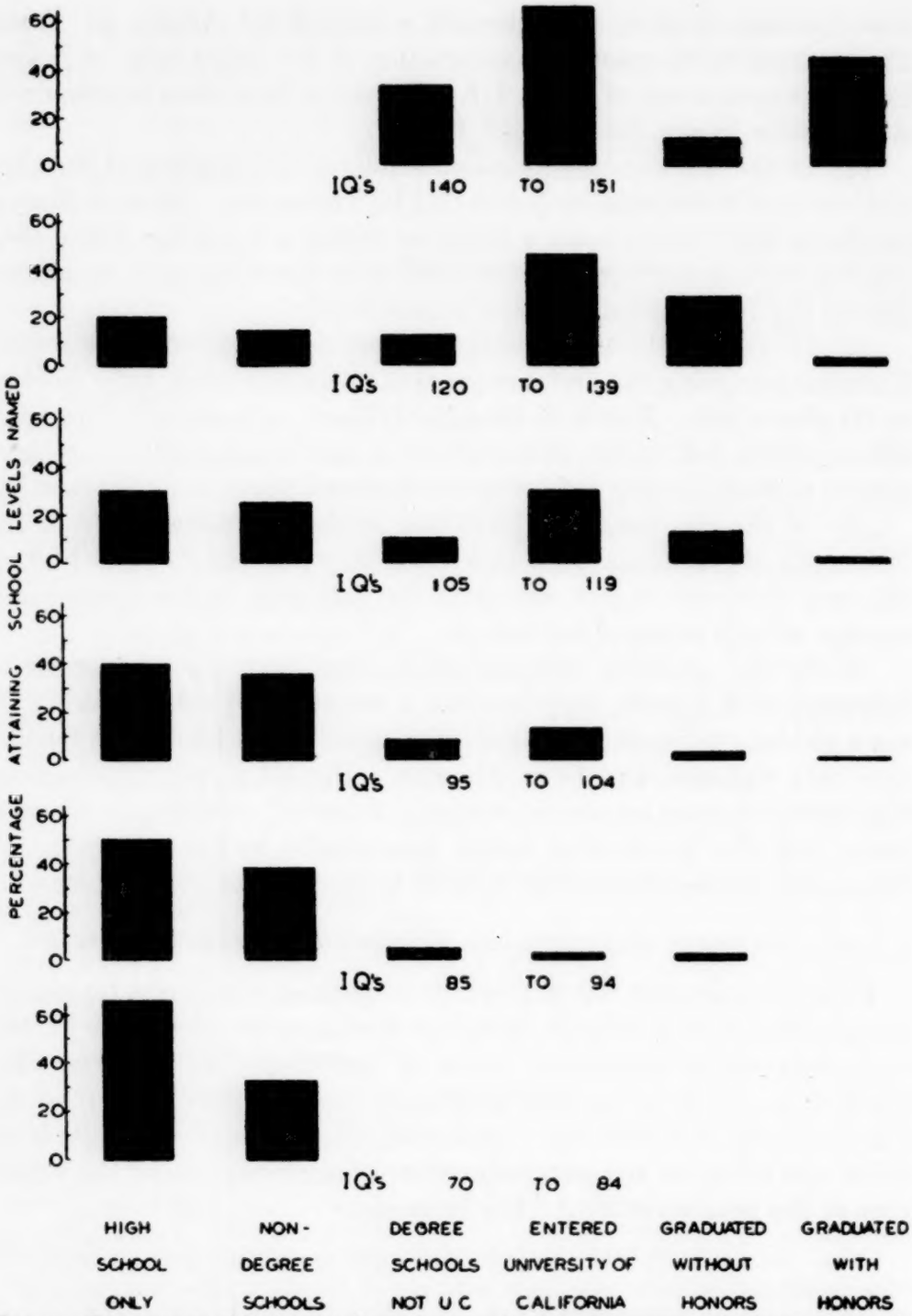

FIG. 1.—Percentage of high-school pupils in each of six IQ groups who later attained to various stages of school progress.

(1) For children of 70 to 84 IQ, the chances are sixty-eight in one hundred, or more than two to one, that schooling will not continue beyond high school, while the probability of entering a degree-granting institution appears to be nil.

TABLE IV.—PERCENTAGE OF HIGH-SCHOOL PUPILS OF PARTICULAR IQ LEVELS WHO ATTAINED TO VARIOUS STAGES OF SCHOOL PROGRESS, INDICATING THE CHANCES IN ONE HUNDRED OF PUPILS OF LIKE INTELLIGENCE DOING THE SAME UNDER SIMILAR CONDITIONS

Intelligence quotients						
70–84 (*N* = 40)	85–94 (*N* = 130)	95–104 (*N* = 280)	105–119 (*N* = 451)	120–139 (*N* = 202)	140–15 (*N* = 19)	
						Progress beyond high school.
67.5	51.5	40.7	30.8	21.3		No schooling beyond high school
32.5	38.5	36.1	25.1	17.8		Transferred to junior colleges, business schools, and other non-degree-granting institutions
	5.4	9.3	12.4	13.4	33.3	Sent transcripts to degree-granting institutions, but did not enter the University of California
	4.6	13.9	31.7	47.5	66.7	Entered the University of California
100.0	100.0	100.0	100.0	100.0	100.0	
						Progress through the University.
	3.1	7.9	10.2	9.4	11.1	Did not obtain junior certificate
		1.4	6.2	5.9		Obtained the junior certificate but no degree
	1.5	4.3	14.0	28.2	11.1	Graduated without honors
		0.4	1.3	4.0	44.4	Graduated with honors

(2) For those of IQ's around 100 (95 to 104), the chances are still forty in one hundred that schooling will cease with high school, while less than one in four is likely to gain admission to any degree-granting institution.

(3) Even for boys and girls of the IQ group from 105 to 119, which includes the average college entrant, the prospects are still that three out of ten will not proceed beyond high school, and scarcely one in three of the Oakland group will enter the University of California.

(4) Among pupils of IQ 120 to 139, however, only about one in five stops short with high school, and nearly half of the group observed actually enrolled at the University of California alone.

(5) The gifted group above 140 shows one hundred per cent applying for admission to some degree-granting institution, and forty-four per cent graduating from the University of California with honors.

(6) From among the four or five per cent of junior-high-school pupils with IQ's of 85 to 94 who later entered the University of California, only one in three succeeded in graduating.

(7) The individual with an IQ in the 120's or 130's has roughly three and one-half times as good a chance (47.5 *vs.* 13.9 per cent) of entering the University of California, and seven times (32.2 *vs.* 4.7 per cent) the likelihood of graduating therefrom as has a youth of 100 IQ.

(8) For students of IQ 140 or higher, the prospects of graduating from the University with honors are one hundred times as great as for the person of average intelligence (95 to 104 IQ), and eleven times as good as for even the 120 to 139 IQ group.

SUMMARY

From eleven hundred twelve boys and girls taking the Terman Group Test under fifteen years of age, the two hundred ninety who later entered the University of California showed a mean IQ of 115.5. Those who withdrew before graduating averaged 112.4; those who graduated, 118.3; and those who graduated with honors, 125.2. The mean IQ's of the groups proceeding to other types of institution are likewise reported in Table II.

As in two long-range studies of group tests in elementary school, the Terman IQ's obtained at junior-high-school level proved only moderately less prognostic of university scholarship than is commonly true of intelligence tests administered on admission to college. Nevertheless, the correlation between these IQ's and the grade-point averages of undergraduates in the University of California was but $.35 \pm .035$. A coefficient of this size would normally be interpreted as too low to be of service for individual prediction. To draw such a conclusion in the present case, however, would be to ignore the circumstance that the correlation is based only on those individuals who actually entered the higher institution. Intelligence quotients below 110 are the rule in high school but the exception in university. The very fact that an individual of IQ 100 or lower has succeeded in gaining admission to an institution of higher learning is

strong evidence that he possesses compensatory qualities of character or other advantages which have enabled him to rise superior to nearly all of his peers in intelligence. The same qualities should serve him well in college work. This tends to produce a deceptively low correlation.

When the chances of attaining a given stage of schooling beyond high school are calculated in terms of the percentage of ninth- or tenth-grade pupils in each IQ group who do so, the relationship between IQ and such success becomes strikingly apparent. In Table IV these probabilities are presented in a form convenient for use in guidance. While the precise figures are conditioned by the circumstances of the particular time and school studied, the situation from which they are derived is believed to be sufficiently typical to make the data of real value to counsellors.

REFERENCES

1. Adams, F. J.: "Predicting high-school and college records from elementary-school test data." *Journal of Educational Psychology*, Vol. XXIX, January, 1938, pp. 56–66.
2. Bell, H., and Proctor, W. M.: "High-School Populations Then and Now—A Sixteen-Year Span." *School Review*, Vol. XLIV, November, 1936, pp. 689–693.
3. Byrns, Ruth and Henmon, V. A. C.: "Long-range prediction of college achievement." *School and Society*, Vol. XLI, June 29, 1935, pp. 877–880.
4. Ellis, R. S.: *Psychology of Individual Differences.* Appleton & Co., 1928, p. 453.
5. Portenier, Lillian: "A twelve-year study of differentiated groups of high-school pupils." *Journal of Educational Psychology*, Vol. XXIX, January, 1938, pp. 1–13.
6. Segel, David: "Prediction of college success." United States Office of Education *Bulletin* 1934, No. 15, p. 69.

INFERRING TYPE OF PUPIL MENTAL ACTIVITY FROM CLASSROOM QUESTIONS ASKED

STEPHEN M. COREY* AND GEORGE L. FAHEY
University of Wisconsin

PROBLEM

This study was undertaken to determine the degree to which judges concur in their estimates of the complexity of a pupil's mental processes when the estimate is based upon a sample of the classroom questions asked by the pupil. Although there is a considerable experimental and statistical literature on classroom questioning activity,[3,4,6,8,11,12,13,15] no study of this aspect of the problem has been published. Most of the references to questioning which appear in the educational textbooks are very general in nature.[1,2,5,7,9,10,14,16,17,18,19] Attention is rarely directed to the kind of questions which pupils ask. The general importance of teacher questioning is stressed, but little is made of those aspects of pupil inquiries which have important educational implications. One frequently hears the assertion that a certain pupil asks "thoughtful" questions but it is assumed that the teacher is in a position to identify such inquiries when they occur.

PROCEDURE

Our procedure involved: (*a*) The selection of a random sampling of the questions asked by ten seventh-grade pupils, (*b*) the duplication of these questions for distribution among judges, (*c*) the evaluation of the questions by the judges, and (*d*) a statistical study of the reliability of the estimates of the judges. An earlier investigation had made available verbatim records of all of the classroom questions asked by thirty-three children during a year in a seventh-grade science class. The records of ten of these pupils were selected for study in accord with the following criteria: (*a*) Each of the ten must have asked at least twenty questions during the course of the year,† (*b*) each sex

* The senior author suggested the problem and shared responsibility for the form of the present report. Both authors are grateful to the Research Committee of the University of Wisconsin's Graduate School for a grant which defrayed in part the cost of a series of investigations of which this study is one.

† The authors were interested to learn that two of the pupils asked no questions before the entire class during a year of science instruction. The median number asked was sixteen. Only twenty-five per cent of the pupils asked more than thirty-eight questions.

must be equally represented, and (*c*) the pupils studied must represent an approximation of the distribution of IQ's in the total class.

The twenty questions asked by each pupil were drawn by chance from the total number he had asked and were mimeographed on one page in random order. This sheet contained no information other than a symbol to identify the pupil. In order that no constant error might occur in the estimates as a result of one paper being first or last or in any other constant position the papers given the various judges were stapled together in all possible sequential arrangements. Each of the twelve judges was given the twenty questions asked by each of ten pupils and the following mimeographed directions:

> Attached hereto are ten pages of sample questions asked by ten seventh-grade pupils during an academic year. These questions and the pupils which they represent have been selected from a class of thirty-three seventh-grade pupils. The only controls used in the selection were the choice of an equal number of both boys and girls and the assurance of a fairly normal distribution of intelligence quotients.
>
> Will you please read the questions on each page critically, and place a check mark in the left-hand margin opposite those questions which you believe imply the operation of the "higher mental processes." A question should be checked if by its nature is implied a clear grasp of a problem, or an attempt at selection of data, or suspended judgment, or other signs of thinking. In addition to checking these best questions will you please encircle the check mark for the question which you consider the best of all on each page.

Approximately forty minutes were devoted to the evaluations. Of the persons who acted as judges, six had earned the Master's degree and three the degree of Doctor of Philosophy. Seven of the judges had had teaching experience and five were receiving graduate training in a department of education. The authors expected numerous questions from the judges as to just what they were to do because the difficulty of identifying with any certainty a "thought" question as opposed to a "fact" question had been discussed. Actually, however, the judges went to work with an attitude that implied they knew what they were looking for. The results below indicated that in a large measure they did.

An analysis of the data was made in such a fashion as to yield answers to the following questions:

(1) How much agreement was there among the various judges in terms of the implications of particular questions as to type of pupil mental activity?

(2) How much agreement was there among the judges in terms of their evaluation of the complete list of questions asked by a particular pupil?

(3) What was the relationship between the ranking assigned the papers by individual judges and by selected groups of judges?

(4) What was the reliability of these rankings?

(5) What was the relationship between the judges' estimates of a pupil based on his classroom questions and: (*a*) the pupil's IQ, (*b*) the final grade given the pupil in the science course, and (*c*) a pupil's interests as indicated by an interest inventory?

RESULTS

The agreement of the judges on individual questions is given in Table I. Two of the two hundred questions were agreed upon by

TABLE I.—NUMBER OF QUESTIONS AGREED UPON BY FROM NO TO TWELVE JUDGES AS IMPLYING A RATHER HIGH TYPE OF MENTAL ACTIVITY

45 questions out of 200 were checked by	no	judges as implying thought
43 questions out of 200 were checked by	one	judge as implying thought
21 questions out of 200 were checked by	two	judges as implying thought
18 questions out of 200 were checked by	three	judges as implying thought
21 questions out of 200 were checked by	four	judges as implying thought
16 questions out of 200 were checked by	five	judges as implying thought
3 questions out of 200 were checked by	six	judges as implying thought
13 questions out of 200 were checked by	seven	judges as implying thought
5 questions out of 200 were checked by	eight	judges as implying thought
7 questions out of 200 were checked by	nine	judges as implying thought
6 questions out of 200 were checked by	ten	judges as implying thought
2 questions out of 200 were checked by	eleven	judges as implying thought
0 questions out of 200 were checked by	twelve	judges as implying thought

eleven of the twelve judges as indicating rather profound thinking on the part of the pupils who asked them. There are given below illustrative questions of various types. In the first group are those which none of the judges thought implied very much, if any, higher mental activity. The second group contains some of the questions upon which half of the judges were in agreement and the third group questions which ten out of the twelve judges believed clearly implied that the pupils who asked them were thinking rather carefully.

Group one: Questions none of the judges checked as implying mental activity of a very high order.

1. Did we see the jack pine?
2. What's the fourteenth question?
3. Will just "kinglet" be enough for 25?

Group two: Questions which half of the judges checked as implying thought on the part of the pupil.

1. In the winter there would be snow on the glass (to shut out the sunlight) wouldn't there?
2. About how long will it take before the eggs will be frogs?
3. I don't see how they can make writing material out of cotton?

Group three: Questions which ten out of twelve judges believed implied very clearly that the pupils asking them were thinking.

1. Why does the water in a syphon stop when it reaches the same level?
2. Will you tell me why, when you take hold of a real strong current, you can't let go?
3. If a man didn't have a nervous system and he was going to be electrocuted, could it kill him?

Not shown in Table I but of considerable interest was the agreement of the judges as to the number of questions checked for each pupil. The range was from one to thirteen—that is, one judge believed that only one out of the twenty questions asked by a particular pupil

TABLE II.—RANKING OF EACH PUPIL IN TERMS OF THE NUMBER OF HIS QUESTIONS WHICH WERE CHECKED BY EACH JUDGE AND BY COMBINATIONS OF JUDGES

Ranked by	Ranks assigned pupils *A*, *B*, *C*, *D*, etc.									
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Judge 1	6.5	6.5	6.5	9.5	9.5	1.5	4	1.5	6.5	3
Judge 2*	8	5.5	2	10	5.5	5.5	9	5.5	2	2
Judge 3	4	7	4	4	10	7	1	2	7	9
Judge 4	1.5	4.5	8	4.5	8	10	8	1.5	4.5	4.5
Judge 5	4	3	9.5	6.5	9.5	6.5	1.5	1.5	6.5	6.5
Judge 6	1	6	6	10	6	3	3	3	8.5	8.5
Judge 7*	4	2	2	9.5	6.5	9.5	6.5	6.5	6.5	2
Judge 8*	7.5	1	10	2.5	2.5	4	9	5.5	7.5	5.5
Judge 9	4	4	7	8.5	10	6	1	2	8.5	4
Judge 10	1.5	3.5	7.5	9.5	9.5	5.5	3.5	1.5	7.5	5.5
Judge 11	9.5	2.5	6.5	4	9.5	6.5	1	2.5	6.5	6.5
Judge 12	4.5	9.5	3	4.5	9.5	7	2	1	7	7
Range of ranks assigned by twelve judges	1–9.5	1–9.5	2–10	4–10	2.5–9.5	1.5–10	1–9	1–6.5	2–8.5	2–9
Rank based on summed estimate of all twelve judges	4	3	6.5	9	10	6.5	2	1	8	5

* Judges 2, 7, and 8 were the "poorest" three judges in the sense that their estimates correlated least with the composite or summation rank.

implied thinking of a very high order while the maximum number of questions checked for any pupil was thirteen. For all ten pupils (two hundred questions) one judge checked seventy-two questions and another as few as twenty-seven in terms of their implications regarding mental activity of a high order. Adding the checks made by all judges, one pupil's questions were checked a total of forty-four times while the questions of another were given seventy-nine checks.

TABLE III.—RAW CORRELATIONS MEASURING RELATIONSHIP BETWEEN ESTIMATES OF INDIVIDUAL JUDGES AND CRITERION ON THE BASIS OF THE NUMBER OF QUESTIONS CHECKED AS IMPLYING HIGHER MENTAL ACTIVITY

	r	PE
Correlation between:		
First six judges and second six judges (chance division)	.87	± .04
Judge 1 and criterion	.67	
Judge 2* and criterion	−.09	
Judge 3 and criterion	.60	
Judge 4 and criterion	.42	
Judge 5 and criterion	.88	
Judge 6 and criterion	.62	
Judge 7* and criterion	.37	
Judge 8* and criterion	−.22	
Judge 9 and criterion	.97	
Judge 10 and criterion	.93	
Judge 11 and criterion	.63	
Judge 12 and criterion	.51	

* Judges 2, 7, and 8 were the three "poorest" judges.

In order that comparisons might be made between individuals each pupil was given a score which represented the number of times his questions were checked by each judge. This made it possible to rank the pupils in terms of their gross scores or in terms of partial scores resulting from different groupings of the twelve judges. The judges were not told that this ranking would be made. The results of the rankings appear in Table II which shows considerable agreement among the individual judges. A large part of the disagreement can be explained in terms of the rankings assigned by the "poorest" three judges. (See note to Table II.)

Since no "true" ranking was available, the composite ranking was assumed to approximate a true ranking more closely than that made as the result of the judgments of any one person. Of the various

possible methods of making a composite ranking the one obtained by using the total number of checks placed on the paper of a pupil by all judges was finally used. This ranking will be referred to hereafter as the criterion ranking. The ranks assigned each paper by each judge were correlated (Spearman rank method) with this composite in order to obtain some idea of the agreement among the judges with the criterion. These correlations are given in Table III and are self-explanatory.

Although the magnitude of the relationships between the estimates of individual judges and the criterion was not great ($Md.r = +.62$), with the exception of judges 2, 7, and 8, they were in an appreciable degree making their evaluations in terms of some common element or elements. The agreement between two chance groups of six judges each was quite high ($r = +.87$).

In order to find out more about the reliability of individual judges than was shown by their correlation with the composite, intercorrelations were computed between all judges. The frequency distribution of these intercorrelations is given in Table IV. All seventeen of the

TABLE IV.—TABLE SHOWING THE DISTRIBUTION OF INTERCORRELATIONS BETWEEN INDIVIDUAL JUDGES
(Rank Method)

COEFFICIENTS	FREQUENCY
0.76– 1.00	5
0.51– 0.75	14
0.26– 0.50	18
0.01– 0.25	12
0.00––0.25	12
−0.06––0.50	4
−0.51––0.75	1
	—
	66

negative correlations involved the three judges who agreed least with the composite ranking. Nor did these three judges agree with one another to an appreciable extent. The highest positive agreement, ($r = +.92$), was found between judges 3 and 12 who were both near the median with respect to agreement with the criterion. The two judges having the highest agreement with the composite, judges 9 and 10, also agreed with one another to the extent of +.90. These latter two judges were involved in most of the high correlations between judges.

From the beginning it was recognized that the judgments regarding the implications of questions might be related to other measures of

pupil achievement or ability. Consequently, the rankings assigned to the questions of individual pupils were compared with the rankings of the same pupils on other variables measured by more or less standardized instruments. These comparisons are listed below. All are computed by the rank method of correlation.

Correlations between intelligence quotients and:

	r
Ranks assigned by "best" six judges	.13
Ranks assigned by all judges	.17
Ranks assigned by "poorest" three judges	−.13

Correlations between grades in the science course and:

	r
Ranks by "best" six judges	−.28
Ranks by all judges	−.40
Ranks by "poorest" three judges	−.51

Correlations between general scholastic average and:

	r
Ranks by "best" six judges	−.11
Ranks by all judges	−.24
Ranks by "poorest" three judges	−.10

Correlations with "Likes" in science as measured by the "Interest Questionnaire" produced by the Progressive Education Association and:

	r
Ranks by "best" six judges	−.38
Ranks by all judges	−.32
Ranks by "poorest" three judges	.12

Correlations with "Dislikes" in science as measured by the "Interest Questionnaire" of the Progressive Education Association and:

	r
Ranks by best six judges	+.01
Ranks by all judges	+.02
Ranks by "poorest" judges	−.13

None of these correlations was particularly significant with the possible exception of those involving course grades. These were consistently negative and the implication is rather disturbing. Considering all of the correlations, their magnitude indicates that the judgments about questions were based upon factors other than those measured by the standardized tests described above.

The possibility that questions may have been checked for some other reason than their thought implications is a valid one. The only

factor of this nature that seemed adapted to objective measurement was the sentence length of questions. The correlation between the length of questions in words and the number of times each question was checked was found to be +.50. This relationship does not necessarily mean that the judgments were irrespective of thought implications, however. Questions limited to less than five words have little room in them to convey much evidence of profoundness of thought and questions of greater than fifteen words obviously necessitate some mental effort in their very formulation.

SUMMARY AND CONCLUSIONS

The purpose of this paper was to discover whether or not judges evidence any agreement in their evaluations of the thought implications of pupils' questions. Two hundred questions, twenty of which were asked by each of ten pupils, were submitted to twelve judges who checked the questions which they considered as indicative of rather profound mental activity on the part of the questioner. Various methods of determining the reliability of judgments were applied. Within the limits of the data, the following inferences seem justified: (*a*) Approximately seventeen per cent of the two hundred questions were agreed on by more than half of the judges as indicative of pupil thinking. (*b*) The agreement of individual judges as to the ranking of the various pupils was marked. (*c*) Several intercorrelations of rankings by groups of judges yielded coefficients clustering around +.85 which coefficient might well represent the reliability of the method of evaluation. (*d*) No significant relationship was found between the combined judges' evaluations of the thought implications of pupils' questions, and such factors as intelligence, grade in course, scholastic average, and interest in science. (*e*) The relationship between length of question in words and number of times checked as indicative of mental complexity was about +.50.

BIBLIOGRAPHY

1. Adams, Jesse E., and Taylor, William S.: *An Introduction to Education and the Teaching Process.* The Macmillan Company, New York, 1932, pp. 207; 222–223.
2. Avent, J. E.: *Excellences and Errors in Teaching Methods.* Jos. E. Avent, Publisher, Knoxville, Tennessee, 1931, p. 253.
3. Dale, Edgar: "Children's Questions as a Source of Curriculum Material." *Educational Research Bulletin,* Vol. XVI, 1937, pp. 57–66.
4. Davis, Edith A.: "The Form and Function of Children's Questions." *Child Development,* Vol. III, 1932, pp. 57–74.

5. Douglass, Harl: *Modern Methods in High-School Teaching.* Houghton Mifflin Company, Boston, 1926, pp. 44–45.
6. Finley, Chas. W.: "Some Studies in Children's Interests in Science Materials." *School Science and Mathematics,* Vol. XXI, 1921, pp. 1–24.
7. Fitzpatrick, F. B.: *Present-Day Standards for Teaching.* F. A. Owens Publishing Company, Dansville, New York, 1926, p. 90.
8. Gatto, Frank M.: *Pupils' Questions: Their Nature and Their Relationship to the Study Process.* Unpublished Ph.D. Dissertation, University of Pittsburgh, 1929. Annotated in: *Annotations of Theses and Dissertations in Education.* University of Pittsburgh, School of Education, Published by Xi Chapter, Phi Delta Kappa, March, 1934, p. 44.
9. Griffith, C. R.: *Psychology Applied to Teaching and Learning.* Farrar and Rinehart, Inc., New York, 1939, pp. 476–477.
10. Hassler, J. O., and Smith, R. R.: *The Teaching of Secondary Mathematics.* The Macmillan Company, New York, 1930, pp. 172; 194–195.
11. Helseth, I. O.: *Children's Thinking: A Study of the Thinking Done by a Group of Grade Children When Encouraged to Ask Questions About United States History.* Teachers College, Columbia University, New York, 1926.
12. Stevens, Romiett: *The Question as a Measure of Efficiency in Instruction.* Teachers College Contributions to Education, Columbia University, New York, 1912.
13. Storm, H. C.: "Who Can Ask the Better Questions—Teachers or Pupils?" *Elementary School Journal,* Vol. XXVIII, 1928, pp. 610–615.
14. Symonds, P. E.: *Education and the Psychology of Thinking.* McGraw-Hill Book Company, Inc., New York, 1936, pp. 46–47.
15. Thompson, Lou H.: "Children's Questions." *Educational Research Bulletin,* College of Education, The State University of Ohio, Columbus, Ohio, Vol. III, 1924, pp. 350–352.
16. Umstattd, J. G.: *Secondary-School Teaching.* Ginn and Company, Boston, 1937, pp. 203; 272–273.
17. Waples, Douglas: *Problems in Classroom Method.* The Macmillan Company, New York, 1927, pp. 316, 348, 446.
18. Wesley, Edgar Bruce: *Teaching the Social Studies.* D. C. Heath & Company, New York, 1937, p. 489.
19. Yoakam, G. A., and Simpson, R. G.: *An Introduction to Teaching and Learning.* The Macmillan Company, New York, 1934, p. 106.

STUDY OF THE RELIABILITY AND VALIDITY OF THE ACADEMIC PREFERENCE BLANK[1]

VINCENT J. SHARKEY AND JACK W. DUNLAP

University of Rochester

The Academic Preference Blank is designed to provide estimates of students' achievement in various school subjects in terms of their expressed preferences for certain things, persons, localities, and activities. There are two forms of the Blank, *A* and *B*, each of which is composed of ninety items.

The value of such an instrument in the estimation and prediction of school success is dependent upon its reliability and validity. Although the items of the Blank were selected on the basis of the constancy and validity as determined by the data upon which the test was constructed, it is necessary that these statistics be investigated in an entirely different sample of the population. It was for this purpose that the studies reported herein were carried out.

The reliability coefficients were determined by the alternate form and the test-retest methods. The validity coefficients were obtained by correlating the differential scores on the Academic Preference Blank with the scores on the various subtests of the New Stanford Achievement Test, the subtests of the Metropolitan Achievement Test, Form *A*, and with the scores on the Terman Group Test of Mental Ability, Form *A*, for each of the grades investigated. The plan of administration of the tests is presented in Table I. In order that the Blank might be tested for the prediction of future achievement the Academic Preference Blank was administered six months previous to the achieve-

TABLE I.—PLAN OF ADMINISTRATION OF THE TESTS

Grade	Total number	Metropolitan	Stanford	Terman	Preference *A vs. B*
IX	289	...	...	258	289
VIII	271	118	153	226	271
VII	300	116	182	267	300
VI	320	129	191	280	320
VII–VIII	126	...	...	...	*A vs. A*
VII–VIII	153	...	...	...	*B vs. B*

[1] This is one of a series of studies from the Department of Education Statistics, University of Rochester.

ment tests. It is thus apparent that the expressed preferences could not have been a function of achievement at the time achievement was measured.

The reliability coefficients for the various grade groups are presented in Table II. When the alternate form method was employed, the reliabilities for the school subjects, MA and IQ for the sixth-grade range from .74 to .85, with a median of .83. For the seventh grade,

TABLE II.—RELIABILITIES OF THE DUNLAP ACADEMIC PREFERENCE BLANK FOR THE VARIOUS GRADE GROUPS

	Alternate form				Test-retest	
Grade	VI	VII	VIII	IX	VII and VIII	
Form	*AB*	*AB*	*AB*	*AB*	*AA*	*BB*
Paragraph meaning	.81	.78	.75	.69	.77	.89
Word meaning	.85	.81	.81	.56	.86	.89
History	.84	.83	.77	.75	.86	.91
Grammar	.78	.76	.69	.75	.77	.90
Geography	.85	.83	.79	.77	.90	.91
Literature	.82	.82	.79	.76	.82	.90
Arithmetic	.74	.75	.76	.67	.74	.84
General achievement	.83	.79	.74	.68	.84	.90
Mental age	.83	.81	.70	.70[1]	.83	.88
IQ	.82	.74	.73	.69[1]	.81	.88
Median reliability	.83	.80	.75	.70	.82	.90
N	320	300	271	114	126	153

[1] $N = 289$.

the reliabilities range from .74 to .83, with a median of .80. For the eighth grade the reliabilities range from .69 to .81, with a median of .75 and for the ninth grade the range is from .56 to .77, with a median of .70.

The test-retest method of determining reliability was used with two other groups composed of seventh- and eighth-grade students. This method disclosed coefficients which range from .74 to .90 (median .82) for Form *A* and from .84 to .91 (median .90) for Form *B*.

Table III contains the median reliabilities of the various subtests of the Stanford, Metropolitan, and Academic Preference Blank. Those for the Metropolitan were derived from data for Grades VI to

IX reported in the manual for that test. Those for the Stanford were published in the manual for the Stanford as average reliabilities for the various subtests, while those for the Preference Blank are the median reliabilities of the various grades investigated.

Thus it can be seen that the Academic Preference Blank compares favorably with two well-known achievement tests as far as the reliability of the subtests is concerned. The only serious discrepancy occurs in the case of the general achievement reliability, and this is to be expected since the total Metropolitan requires four and one-half hours of administration time, the Stanford requires three hours of administration time, while the Preference Blank may be administered in about fifteen minutes, or a ratio of 18:12:1.

TABLE III.—MEDIAN RELIABILITIES OF THE VARIOUS SUBTESTS OF THE METROPOLITAN, STANFORD, AND ACADEMIC PREFERENCE BLANK

Subtest	Metropolitan	Stanford	Preference blank
Paragraph meaning	.84	.85	.78
Word meaning	.87	.86	.81
History	.82	.74	.83
Grammar	.84	.75	.76
Geography	.81	.85	.83
Literature	.86	.68	.82
Arithmetic	.89	.83	.75
General achievement	.92	.94	.79
Average *N*	193	197	300

From these data it may be safely concluded that the Academic Preference Blank is sufficiently reliable to justify its usage in the estimation of achievement in the school situation providing the validity coefficients warrant such application. It should be particularly useful for those situations where one is concerned with comparing total groups such as classes, grades, schools, or cities. Hence, it is upon the validity coefficients that we next focus our consideration.

The validity coefficients for the Academic Preference Blank for the various grades and for the three criterion tests employed are presented in Tables IV and V. When the several scores on the Academic Preference Blank were correlated with the scores on the corresponding subtests of the Metropolitan, the validities ranged from .32 to .73, with a median of .57. When the subtests of the Stanford were used as

criteria, the validity coefficients ranged from .28 to .72, with a median of .52. While it is true that some of these values are not great enough to reduce appreciably the standard error of estimate upon which

TABLE IV.—VALIDITY OF THE DUNLAP ACADEMIC PREFERENCE BLANK FOR VARIOUS GRADES AND VARIOUS CRITERIA

	Criterion—subtests of Metropolitan						Criterion—subtests of New Stanford					
Grade	VI		VII		VIII		VI		VII		VIII	
Form	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
Paragraph meaning	.68	.65	.67	.55	.53	.58	.63	.59	.66	.57	.46	.37
Word meaning	.61	.63	.66	.61	.57	.47	.71	.64	.70	.61	.49	.41
History	.68	.64	.55	.53	.56	.52	.63	.53	.59	.58	.47	.46
Grammar	.59	.59	.61	.64	.39	.45	.52	.52	.54	.51	.28	.34
Geography	.51	.52	.49	.45	.56	.46	.47	.40	.64	.62	.44	.45
Literature	.73	.69	.60	.58	.41	.36	.60	.58	.60	.55	.47	.43
Arithmetic	.59	.49	.50	.43	.32	.39	.40	.41	.43	.41	.41	.33
General achievement	.65	.68	.65	.60	.57	.56	.72	.68	.69	.67	.50	.54
N	129	129	116	116	118	118	191	191	182	182	153	153

predictions may be made, nevertheless, in general, they are considerably higher than those which have been previously reported for measures of interests.

TABLE V.—VALIDITY OF THE DUNLAP ACADEMIC PREFERENCE BLANK FOR VARIOUS GRADES

	Criterion—Terman group test of mental ability							
Grade	VI		VII		VIII		IX	
Form	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
Mental age	.70	.67	.68	.62	.63	.56	.64	.49
IQ	.62	.60	.66	.63	.55	.50	.58	.40
N	280	280	267	267	226	226	258	258

The intercorrelations between the subtests of the Stanford and Metropolitan Achievement Tests were available for four hundred thirty-seven seventh- and eighth-grade students. The median of

the intercorrelations for the subtests for which an interest can be scored is .75. This value is probably slightly higher than would have been obtained if a single grade range had been employed. The index of forecasting efficiency for predicting one achievement test from the knowledge of the other is .333. The median correlation between the subtests of the Stanford and Form *A* and Form *B* of the Academic Preference Blank is .60. An identical value was found for the median correlation between the scores on the Academic Preference Blank and the subtests of the Metropolitan. The index of forecasting efficiency of the preference blank is .20 in both cases. Thus the preference blank is sixty-eight per cent as efficiency in predicting achievement as one achievement test is in predicting the other test. The value is probably low owing to the dilation of the correlation between the Stanford and Metropolitan due to the use of a two-grade range.

Table V contains the correlations between the Academic Preference Blank and the Terman Group Test of Mental Ability for the different grade groups. As can be seen, these range from .40 to .70, with a median of .62. As heretofore, these are also higher than previously reported coefficients and, while their predictive efficiency is not too great, the indications are that the Academic Preference Blank may be employed to obtain a rough estimation of MA or IQ in a short period of time. Pintner has collated a number of studies on the correlation between various group tests of mental ability and the median r is only .70.[1] Thus the median r between the Terman test and the estimates of mental ability obtained from the Academic Preference Blank is not much lower than the r between two group intelligence tests.

It may well be that the correlation between the preference blank and achievement is due to the saturation of the preference blank with the factor of intelligence. If this is the case, the blank is nothing but a disguised intelligence test. In an attempt to answer this question, two analyses were made. Consider two individuals who have equal intelligence quotients, but who radically differ in achievement; obviously, some factor besides intelligence must account for the difference. Thus, if two groups of individuals were selected so that they were paired as to IQ but differ greatly in achievement, the preference factor, if it is a reality, should appear in terms of widely different preference scores for the two groups.

---

[1] Pintner, R. A.: *Intelligence Testing*, 1930, p. 57.

The two groups were selected from the sixth-grade population matched as to IQ but different in achievement as measured by the various subtests. An individual was considered matched with another individual with regard to IQ if their IQ scores differed by not more than two points. The same individuals were considered different as to achievement on the subtest under consideration if their scores differed by at least ten points, except in the cases of the Stanford Arithmetic subtest and the total achievement scores. Since the Stanford Arithmetic score was the sum of the scores on two subtests, the criterion for difference on this variable was at least twenty points. Since the

TABLE VI.—CRITICAL RATIOS OF THE DIFFERENCE BETWEEN THE MEAN PREFERENCE BLANK SCORES OF TWO GROUPS MATCHED AS TO IQ BUT DIFFERENT AS TO ACHIEVEMENT SCORES ON THE VARIOUS SUBTESTS
(Sixth Grade)

	Paragraph meaning	Word meaning	History	Grammar	Geography	Literature	Arithmetic	General achievement
Form *A*.								
*M*1–*M*2	5.98	10.99	12.41	3.26	4.34	9.59	1.86	12.82
$\sigma_{diff.}$	1.94	2.52	2.16	1.58	2.15	2.06	1.23	2.71
CR	3.08	4.36	5.75	2.06	2.02	4.66	1.51	4.73
Form *B*.								
*M*1–*M*2	7.35	10.82	11.70	5.24	4.31	9.34	2.16	14.30
$\sigma_{diff.}$	2.26	2.57	2.20	1.52	2.23	1.99	1.20	2.49
CR	3.25	4.21	5.32	3.45	1.93	4.69	1.80	5.74
N	90	89	91	101	105	90	112	72

general achievement or total score on the Stanford was the sum of the scores of ten subtests and on the Metropolitan, the sum of the scores of seven subtests, the criterion for difference on this variable was one hundred for the former and seventy for the latter. These differences were approximately 1 SD of the distributions, which were one hundred seven and seventy-six, respectively.

The critical ratios of the differences between the mean preference blank scores for the groups matched as to IQ but different as to the various achievement subtest scores are presented in Table VI. All but five of these critical ratios meet the criterion for statistical significance. The differences that are not significant all run in the expected

direction and in three of these cases the chances are at least ninety-seven in one hundred that the difference is significant. It would seem, therefore, that whatever factor is causing the differences in achievement when intelligence is held constant, is also operating to cause differential scores on the Academic Preference Blank.

The converse of this situation should hold; namely, that if two groups of individuals were matched both as to achievement and to IQ, there should be no difference in their mean preference scores. The criterion for matching on IQ was the same as that used above. Two individuals were considered matched as to achievement if their scores on the subtests differed by not more than two points, except in

TABLE VII.—CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE MEAN PREFERENCE BLANK SCORES OF TWO GROUPS MATCHED AS TO IQ AND ACHIEVEMENT SCORES ON THE VARIOUS SUBTESTS
(Sixth Grade)

	Paragraph meaning	Word meaning	History	Grammar	Geography	Literature	Arithmetic	General achievement
Form *A*.								
*M*1–*M*2	3.60	2.10	0.40	0.20	1.66	0.60	0.49	1.21
$\sigma_{diff.}$	2.15	2.47	1.95	1.62	2.28	2.15	1.42	2.05
CR	1.67	0.85	0.21	0.12	0.73	0.28	0.35	0.59
Form *B*.								
*M*1–*M*2	3.40	0.59	0.52	1.01	3.51	2.68	0.65	1.78
$\sigma_{diff.}$	2.50	2.49	2.08	1.59	2.37	2.40	1.42	1.79
CR	1.36	0.34	0.25	0.64	1.48	1.11	0.46	0.99
N	71	87	94	87	83	70	78	102

the cases of the Stanford Arithmetic subtest and the total scores on the Stanford and Metropolitan tests. For the reasons noted above, the criterion for matching on the Stanford Arithmetic was four points; on the total Stanford twenty points; and on the total Metropolitan fourteen points. All of these maximum differences are approximately two-tenths of a standard deviation of the distribution scores for each test.

The critical ratios between the mean preference blank scores for the matched groups for each of the variables are presented in Table VII. All of these differences are thus seen to be insignificant. The largest critical ratio is 1.67 and a great majority of the rest are below 1.00.

Thus, two groups of individuals matched as to intelligence and achievement are also equivalent, within sampling errors, in their preference scores.

In conclusion, it would appear that the Academic Preference Blank has substantial reliability for each of the ten factors it is designed to measure. The validity of the blank is sufficient to warrant its use when a survey of a group is desired. The evidence indicates that although the blank is partially saturated with the factor of intelligence, in addition it is measuring another factor contributing to achievement. Whether this factor is interest, preference, or what, is beyond the scope of this study.

A CRITICAL ANALYSIS OF VARIOUS METHODS OF CLASSIFYING SPELLING ERRORS, I*

GEORGE SPACHE

Friends Seminary, New York City

In the hope of delimiting a technique for classifying spelling errors, the writer will review a number of proposed systems of classification. Each system will be described and evaluated in terms of its objectivity, ease and clarity of administration, and its value for the diagnostician. In order that their evolution may be more apparent, the studies are described in chronological order from the earliest in 1900 to the present.

The earliest system for classifying spelling errors, in any of the bibliographies of studies related to spelling used by the writer[3,14,20,37,44] is that of Douse in 1900. He noted the following errors:

Interference.
  Prolepsis—assimilation from ahead as spooped for stooped.
  Metapedesis—overleaping, as preced for preceded, femine for feminine.
  Metallage—cross compensations as silibants for sibilants.
  Opisthomimesis—assimilation from the rear as synonyns for synonyms.
Contamination.
  Association with some other word as Tuetonic by association with Tuesday.

It is not necessary to explain Douse's terms further for the present since the reader will note that they are amply defined and illustrated in the work of later writers. Suffice it to say that, "Metapedesis" has since come to be known by the simpler term—omissions and that "Metallage" is better known by the term—partial reversals. Aside from the complexity of nomenclature, Douse's systems is characterized by its subjectivity and paucity of error types. Casual inspection of almost any child's spelling paper reveals that omissions, additions, substitutions, and transpositions comprise the majority, if not all, of the spelling errors. These four types, at least, are requisite for analysis. Studies of later writers will indicate that such categories as overleaping and contamination are in need of more careful definition.

Cornman's system of classification[9] has been adequately described and used by Kline[21] who lists the following errors:

---

* This is the first of three articles on diagnostic work in spelling. The second will describe the steps in establishing the validity and reliability of the system for classifying spelling errors offered in this article. A third article, "Characteristic errors of good and poor spellers," will appear in the *Journal of Educational Research.*

Motor errors.
  Simple—omission, addition, interchange of m and n.
  Suggestion—transposition.
    wrong letter doubled.
    sensori-motor attraction—repetition of a prominent letter or insertion of a letter or syllable involved in a previously written word.
    ideo-motor attraction—substitution of a letter or group of letters from a somewhat similar word or from an irrelevant associated word.
Sensory errors.
  Phonetic.
  Confusion of letters or syllables—ie for ei, able for ible.
    special cases of confusion.
    doubling and non-doubling.

Although more detailed than Douse's, Kline's (or Cornman's) system is similarly subjective, particularly the sensori-motor and ideo-motor types. These errors undoubtedly do occur, but they must be more carefully defined than these writers do if the spelling analysis is to be applied with equal accuracy by every user. Cornman has made the distinct contribution, however, of recognizing the presence of phonetic errors even though he has not defined them.

Cornman's system was employed by Hazel Floyd[13] in her study of cases of spelling disability and by Wycoff[45] in his study of spelling errors of high-school seniors. In her adaptation of the Cornman system, Floyd made several additions. The additional types of errors that she noted or those she more clearly illustrated were:

Drawing phonetic conclusions.
Omissions—silent letters, other letters.
Confusing alternatives—or, ar, er, ir, ur; tion, sion, shion; s, c, z; i, e, a.
Addition—insertion, final.
Unfinished words.
Words misunderstood and wrong cue.
Pupil's mispronunciation.
Confusing similar graphic responses.
Repetition of dominant elements.
Unclassified and freak errors.

Unfortunately, Floyd fails to distinguish clearly between these types. To this writer, incorrect doubling, addition of the insertion type, and repetition of dominant elements are indistinguishable. This overlapping of types is present in Floyd's application of the system in which she lists errors in two or three categories. The type "phonetic

conclusions" is certainly in need of further division since the work of later writers[7,25,26] indicates that fifty to seventy-five per cent of spelling errors are of a phonetic nature. An excellent addition to classification of spelling errors is the noting of the type due to the pupil's mispronunciation. These Floyd would determine by having the pupil read aloud each of the words misspelled and noting the similarity between mispronunciation and misspelling.

In his doctoral thesis, F. L. Wells[42] noted the following motor or graphic lapses:

Associo-suggestive assimilation, substitutions or omissions as "of men of" for "of men are"; "one of my one" for "one of my own."
Regressive assimilation—as "bug" for "begun on"; "host" for "historical."
Anticipatory transposition—as "Mollis, Ell and Kraft-Ebing" for "Moll, Ellis and Kraft-Ebing"; "maglignant" for "malignant."
Omissions, substitutions, spelling by incorrect analogy.
Writing—"and god" for "and got" (crossing the d).
"say" for "saw" (w with descending loop).
dwarfing of a letter.

The similarity between the classifications of Cornman and Wells is apparent. Wells adds to our list the errors resulting from incorrect analogy, poor penmanship, anticipation, and regression. Most of these would benefit from more objective description.

Foster's[15] eleven types of errors were:

Omissions and additions of silent letters.
Confusion of al, le; of ant, ent; of se, ce, ze; of able, ible or of ance, ence.
Mispronunciation.
Alternatives—ie and ei.
Spelling er as pronounced.
Miscellaneous.

Several of Foster's types might well be combined and others enlarged. There appears to be no objection to combining the four types of confusions and the "spelling er as pronounced" into types called phonetic substitutions for vowel or consonant. Similarly, one may supplement the "omissions of silent letters, additions of silent letters" by adding omissions and additions of sounded letters respectively.

In 1910, Gill[17] noted the following errors among children of an English school:

Omission—silent letter, sounded letter.
Insertion—silent letter, sounded letter.
Miscellaneous—confusion of vowels; inversion of letters; others.

Gill has made the obvious additions to Foster's types of errors. However, he has omitted substitutions (or confusions) of consonants which would seem a logical addition to his "confusion of vowels."

In effect, Rusk's[30] classification of spelling errors represents a step backward toward greater subjectivity. The need for delimiting types is apparent from his list.

Defective capacity for expression—interference of fatigue, excitement, etc. with spelling of known words.
1. Words not completed, as an for and.
2. Words wrongly completed where there is a substitution of part of another word as was for way, there for their.
3. Substitutions of words having little or no apparent connection, as you for who.

Faulty apprehension—caused by insufficient analysis of word.
1. Errors of pronunciation (examiner or pupil).
2. Phonetic spelling of irregular words, as oposed for opposed.
3. Errors in visual analysis, as ie for ei.

In applying Rusk's types, Stead[36] made several additions. To the former's "Defective Capacity for Expression," Stead added:

1. Lack of knowledge of word, as corse for coarse.
2. Use of homonyms, as two for too.
3. Fatigue—letter omitted at end or elsewhere; letter added; words ended or begun wrongly.

To the group "Faulty Apprehension," Stead added faulty visual analysis, *i.e.* letters misplaced, letters doubled.

To Rusk and Stead, fatigue would seem an adequate explanation of almost any type of error. Aside from their explanation of cause, the classification may be criticized for overlapping or the lack of distinction between phonetic spelling and words not completed or wrongly completed. One would also question the validity of the category "lack of knowledge of word" unless this is intended to mean lack of knowledge of meaning of word.

In 1914, Cook and O'Shea[8] noted a number of spelling errors. Their list was:

Chance errors—slips of the pen, as thorm for thorn.
misunderstanding as neverous for nervous.
Inversions, repetitions, mispronunciation.
Lapses—slips, characteristic or idiosyncratic errors.
Phonetic analogy—as pearcing for piercing.
Obscure or elided vowel—tendency to interchange e and a or to substitute e for either i or a.

Doubling of letters—single for double or *vice versa*.
Confusion of letters—confusing letters above and below line.

Although they do occur, classification of spelling errors due to slips or lapses becomes a highly subjective matter. The same is true of errors due to misunderstanding. Errors due to phonetic analogy should be more carefully subdivided than here.

Hollingworth's study of special disability in spelling[19] led her to place emphasis upon lapses as causes of spelling errors. She noted the following types of lapses:

1. Copying the ending of the word just above, as closet, clockt.
2. Including a syllable of a word to follow, as postard card.
3. The tendency to omit letters requiring similar motor responses, as sd for sad, gld for glad.
4. The tendency to write letters with similar kinesthetic elements instead of the correct letter, as dod for dog, forn for form.
5. The substitution of a letter with common visual elements, as goiny for going, storeheeper for storekeeper.
6. The substitution of a letter with common phonetic elements, as celect for select.
7. The transposition of two adjacent letters, as Indain for Indian.
8. The perseveration of an element, especially a dominant element, as the theeth for the teeth.
9. The tendency to omit the final letter when the initial letter of the next word has the same sound, as advise to for advised to.
10. Doubling wrong letter, as frezze for freeze.
11. Idiosyncratic errors—constant similar errors peculiar to the subject.

Hollingworth's types include several distinct from those of her predecessors. Unfortunately, many of the types do not lend themselves to easy recognition. This is particularly true of the omissions because of similar motor response, substitution because of similar kinesthetic or visual elements, and the omission of a final letter because of phonetic similarity with the initial letter of the next word. The validity of these categories might be improved if the examiner were able to have the subject respell orally the misspelled words. The words spelled correctly orally, but misspelled when written, might be attributed to lapses and their types noted.

Speer[34] followed the suggestions of Cook and O'Shea[8] and Hollingworth[19] in her classification of spelling errors. She noted the following in addition to the types of her mentors:

Kinesthetic inertia—repetition because of inability to stop.
Influence of foreign or colloquial pronunciation.

Errors on the silent w, omissions of silent letters.
Confused image of word.
Poor penmanship.

Again one finds spelling errors classified according to factors which undoubtedly influence their nature, but are extremely difficult to isolate in any group of misspelled words. Foreign or colloquial pronunciation and confused imagery do influence spelling but valid techniques must be devised for detecting the extent of their influence in a particular word. Pronunciation may be checked by having the child say the word, but confused imagery would seem too difficult to determine readily, and of little value except with respect to a single word when established.

Brandenburg's types of errors[2] were quite similar to a number of the systems preceding his. Two of his types represent contributions to our cumulative classification; namely, the confusion of consonants having the same sound and the phonetic substitution of part of another word.

Elsie Murray's classification[27] adds another type of error to the total lists; namely, phonetic substitution for a diphthong. It is apparent that we might add a similar category to errors of omission, *i.e.* omission of the subordinate or silent vowel of a diphthong. To omissions of sounded letters, we may add the omission of the major or sounded vowel of a diphthong. Murray's list differs from those of her predecessors in the following:

Phonetic substitutions—of vowels, as a for e, e for i, a for i.
of diphthongs, as i for ei or vice versa.
Spelling by analogy—alright (always) preceeded (proceeded).
Influence of other syllables in same word as neucleus, heriditary.

It would be rather difficult to define the spelling by analogy and the influence of other syllables so as to remove the large element of subjectivity. For this reason, the writer does not feel that they represent a contribution toward clarity of classification.

In 1921, Burt[5] distinguished misspellings from the standpoint of errors and from the standpoint of the pupil. Because his list more clearly defines the various errors he tabulated, it is reproduced in full:

From the standpoint of errors.

1. Substitutions—incorrect letter resembles the correct one in form, sound or in movements necessary for writing.
2. Omissions—one or two letters made by similar movements or resembling each other in sound is omitted.
3. Insertions—letters added because of similarity in movement, form or sound.

4. Transposition of letters or syllables.
5. Duplication and repetition—of a single letter or number of letters.
6. Anticipation and perseveration—a letter is moved forward in the word or one already used is moved backward.
7. Articulatory—the incorrect letter represents the incorrect pronunciation.

From the standpoint of the pupil.

1. Lapses—child knows correct spelling but momentarily unable to give it. These errors as a rule are not repeated.
2. Inventions—child does not know the word and invents a spelling.
3. Habitual errors—here the child has learned an incorrect spelling.
4. Idiosyncrasies—here the child makes some characteristic error as adding an extra letter.

Unless further subdivided Burt's list would not permit distinction between phonetic or non-phonetic errors or those due to similar graphic responses, in substitutions, omissions, and insertions. The necessity for further definition in order to obtain objectivity in such errors as anticipation and perseveration, lapses, inventions and habitual errors has been noted.

In 1922 Lester[23] noted ten types of errors due to lapses. The source of these was Hollingworth's list. In his "Spelling Review,"[24] Lester attributes spelling errors of high-school seniors to the following:

Compounds; possessives; apostrophe; capitals.
Prefixes and suffixes; Latin roots.
Confusion of similar words; mispronunciation.
Final consonant before a suffix; final e before a suffix; final y.
Confusion of ence, ent; ei and ie.

The majority of these are breaches of spelling rules or conventions rather than spelling errors in the sense in which the term has been employed here.

Writing in 1924, Randall[29] noted the following errors:

Errors of ignorance which include peculiar forms of words.
Errors due to careless teaching or lack of attention.
Errors due to defective vision or hearing.
Errors of carelessness due to fatigue.
Poor pronunciation, divided attention, or confusion with similarly sounding words.

In attempting diagnosis of the difficulties of a class, Sister Vigilia[39] listed these errors:

Additions or omissions due to haste in writing, as answere, enter(ed).
Substitutions or omissions due to trying to spell word as it sounds; or due to lack of knowledge or familiarity with word.

Substitutions, omissions or repetitions due to poor visual image.
Substitutions caused by faulty pronunciation, as surprice for surprise.
Errors due to handwriting.

Ina Hill divided her analysis of the spelling errors of a few seventh-grade children into these types:[18]

Perception.	Spelling.	Writing.
Knowledge of word	Omissions	Omissions
Meaning of word	Transpositions	Carelessness
Pronunciation	Substitutions	
Hearing	Additions	
Visual image	Repetitions	
	Irregularities	

In Randall's, Sister Vigilia's, and Hill's lists there is a confusing of error types and probable causes of misspelling. The difficulty of attempting a tabulation of errors and causes simultaneously is obvious. A diagnosis must be deductive in that interpretation of the probable reasons for the errors is derived from the tabulation of the errors themselves. The types of errors must be sufficiently defined and detailed as to prevent overlapping and confusion on the part of the examiner. At the same time they should lend themselves to analytic grouping so that the examiner may readily determine the skills and knowledges as well as the significant habits of the pupil. The error types noted by these three writers do not, on the whole, meet these criteria.

Pupils' difficulties with spelling were noted by the teachers of Richmond, Indiana, under the direction of Georgia Davis.[10] The items of the list they compiled which are pertinent to this discussion are:

1. Has not mastered the steps in learning to spell a word.
2. Writes poorly.
3. Cannot pronounce the words being studied.
5. Does not associate the sound of the letters or syllables with the spelling of the word.
8. Has speech defect.
10. Interchanges letters.
13. Has poor hearing.
14. Writes so poorly that he cannot keep up with the class.
16. Has poor vision.
18. Does not know meanings of words.
19. Adds extra letters.
20. Misses syllables in middle of words.
22. Leaves off last letter.

While this partial list is not suitable in toto for a classification of spelling errors, a number of its items are applicable, *i.e.* 10, 19, 20 and 22. Other difficulties that may be translated in terms of errors are 2 and 5. The careful diagnostician would attempt to determine the relative influence of 1, 3, 8, 13, 14, 16 and 18.

In Master's thesis[25] these errors were tabulated:

Omission of letters other than failure to double.
  Vowels.
  Consonants.
Insertions of letters other than doubling.
  Vowels.
  Consonants.
Inversions.
  Adjacent letters.
  Letters adjacent but one.
Doubling.
  Double for single—vowels, consonants.
  Single for double—vowels, consonants.
Confusions and substitutions.
  Vowels.
  Consonants.
Spelling of wrong word, compounds and hyphenates.
Capitalization, punctuation (apostrophe).
Miscellaneous.

There is no distinction between omission or insertion of single letters and syllables; between silent and sounded letters; or between phonetic and non-phonetic substitutions in this otherwise careful classification.

In her study of the influence of phonics upon spelling, Ethel Bugbee[4] classified spelling errors in terms of their relationship to phonics.

I. Caused by lack of application of knowledge of phonics.
  1. Vowel—one vowel sound for another.
  2. R-affected sounds—as pork for park.
  3. Hearing—lack of ear training as tweny for twenty.
  4. Sound—little or no idea of sound of letters or sounds not associated with letters.
  5. Confusion—digraph, vowel or consonant, one for another.
  6. Principle—does not apply phonic principle.
  7. No vowel—omission of vowel in syllable.
  8. Name—confusing name and sound of letter as n for en.
  9. Consonant—lack of knowledge of sound as l for k.

II. Knowledge of phonics a hindrance.
   1. Two principles as saye for say.
   2. Choice of two principles possible.
   3. Application—principle correct but word non-phonetic.

Bugbee's list also includes a large number of errors characterized as miscellaneous. New types are tense, as work for worked; trying to remember word as seen, as eght for eight; number or person of verb; unanalyzable.

The validity of a classification based entirely upon phonics rests upon the validity of the principles of the phonic system employed and, incidentally, upon the extent to which English is truly a phonetic language. It is not appropriate to enter into a discussion of the relative merits of various systems of phonics or whether or not English is phonetic. Bugbee's classification is noted solely because of its originality. It is apparent that a number of categories overlap and that others are highly subjective. As a method of classifying spelling errors it may have value to the teacher of phonics or, specifically, in the first three grades. Whether it is applicable to the misspellings of older children is doubtful.

In 1928, Welch[41] tabulated the spelling errors of four hundred seventeen high-school seniors in California. He gave considerable attention to confused endings or alternatives. In this respect, Welch's list resembles that of Foster. None of his error types are sufficiently different to warrant reiteration.

Carle[6] compared the errors of freshmen and upper classmen in the University of Kansas in 1928. Like Welch and Foster, she listed numerous types of confusions. Two of her types represent departures from the classifications of others; namely, (1) insertion of silent letters as cheereful and (2) wrong division at the end of a line as corr-ecting. She does not describe how the reader is to determine whether an insertion is that of a silent or sounded letter. The second error type is not, strictly speaking, a misspelling.

Since Book and Harter's study of spelling errors[1] is frequently cited as an exemplary method of classification, it is reproduced here in full:

Errors due to inadequate mental control.
   1. Omission—of a letter or letters as the for they.
   2. Anticipation—the tendency to anticipate a letter that comes later in the word as converstation for conversation.
   3. Repeating or adding a letter as theeth for teeth.
   4. Transposition—as esaily for easily.
   5. Carelessness—as m for n; i for e; surily for surely.

6. Doubling wrong letter—as speel for spell.
7. Interference—part of word misspelled because of sound or form association with preceding word.
8. Forgetting—forgetting part of word and substituting phonetically or in meaning, as author for Arthur.
9. Substitution as dod for dog.

Errors in words not learned.

1. Phonetic spelling as Wensday.
2. Mispronunciation as chimley for chimney; bringin for bringing.
3. Alternatives—confusion of ie and ei; tion, cion; ly, y; able, ible.
4. Doubling as Hellen for Helen.
5. Non-doubling, as galons for gallons.
6. Substitution of similar letters or syllables because of common visual elements as storeheeper for storekeeper, goiny for going.
7. Homonyms—as bare for bear, their for there.
8. Ignorance of word—an attempt to spell by sound conditioned by ignorance.
9. Failure to hear or perceive correctly as bureau for mural.

These comments may be offered with respect to Book and Harter's system of classification:

(1) There is no distinction between phonetic and non-phonetic errors or between errors involving a single letter or a syllable in the categories omissions, repeating or adding, substitutions, transpositions, forgetting and ignorance. A careful classification must take this tendency into account.

(2) The classification of such errors as those called carelessness, interference, forgetting, ignorance, failure to hear or perceive, anticipation and substitution because of similar visual elements becomes a matter of subjective opinion and guesswork unless more carefully defined than here. There is corresponding loss to the diagnostician as such guesswork appears.

(3) The lack of objectivity in such categories as substitution of similar letters because of common visual elements is apparent in the classification of such errors as goiny for going. Hollingworth and Book and Harter would attribute this error to such a substitution. Floyd, following Cornman, would call it confusing similar graphic elements. Cook and O'Shea would probably call it a slip of the pen. Masters or Bugbee would term it a simple substitution.

(4) Several of the types of errors overlap, as alternatives may be phonetic transpositions; mispronunciations are chiefly attempts at phonetic spelling (of mispronounced words), etc.

(5) Several of the types are objective and readily distinguishable; namely, doubling, non-doubling, and homonyms.

In her case studies in spelling, Bernice Leland[22] noted several original types of errors; namely, long pauses, no separation of words, errors due to confusion in word meanings, and random performance as Crimster for Christmas.

Carroll's study of bright and dull children[7] used the following types:

Words omitted—child absent, spelling not attempted.
Addition—single, doubling, others; groups of letters.
Omission—single, doubling, others; groups of letters.
Substitution—single, groups, entire word—actual or irrelevant, transpositions.

It is apparent that no account has been taken of the tendency toward phonetic spelling in Carroll's system. For example, the omission of silent letters, the subordinate vowel of a diphthong, or of a sounded letter are not differentiated under errors of omission. Carroll takes no cognizance of errors that may result from handwriting, confusion of homonyms, carelessness, or lapses.

Umberhine[38] followed Carroll's system in her study of the spelling errors of good and poor spellers. To the errors of addition, she added the addition of an apostrophe. To errors of omission, she added the omission of an apostrophe and of a word by reason of lack of attempt or partial completion. To errors of substitution, she added the use of a lower case letter for a capital, and *vice versa*, and the use of groups of letters unrecognizable as a word by reason of four or more errors.

Mendenhall[26] used several systems of classification. In tabulating his collection of spelling errors, he first divided all errors among the types of omission, addition, substitution, and transposition. Second, all errors were classified as phonetic or non-phonetic. Third, all errors were classified as substitutions of real words or non-words. Finally, all errors were divided into homonyms or non-homonyms. Aside from the duplication of effort in scanning the errors four times, Mendenhall's system is much too general to give adequate information to the examiner.

At first reading, Sartorius's system appears to be more adequate than many of those preceding. Further inspection reveals that it is unnecessarily detailed in some categories and lacking in others. Of the fifteen types of errors enumerated, several may be combined without loss of accuracy of diagnosis. Items 5 and 8 may be combined into a type named "omitting of a silent letter"; items 3, 6 and 9 may be combined into "omitting the subordinate vowel of a diphthong"; item 13 may be included in "omitting a silent letter" or "omitting a syllable."

1. Doubling a consonant.
2. Omitting the second of a double consonant.
3. Omitting the second of a double vowel.
4. Omitting a final e.

5. Omitting a silent letter not a final (exclusive of 4, 6, 10, 13).
6. Omitting the i of the phonogram ai as in nail.
7. Substituting i or ey for final y.
8. Omitting the c of the phonogram ck.
9. Omitting the a of the phonogram ea as in each.
10. Transposing final el making it le.
11. Substituting hard k for hard c.
12. Substituting s for soft c.
13. Omitting the e or ed in words ending in ed.
14. Interchanging ei and ie.
15. Substituting a vowel for a vowel.

In Wells's study of the effectiveness of specific remedial procedures in spelling,[43] he noted the following types of errors:

1. Incorrect phonetic spellings.
2. Omissions of letters, of words, omissions to shorten word as flit-flight.
3. Confusing m and n, or alternatives.
4. Additions of letters, transpositions of letters.
5. Unfinished words as sid for side, grap for grape.
6. Words not recognized as uer for other.
7. Similar graphic responses or repetition of dominant element.
8. Mispronunciation, freak errors, poor penmanship.
9. Inattention to first or latter part of word, as rly for early, etc.
10. Wrong word spelled correctly; wrong word incorrectly spelled.
11. Failure to double.

Although Wells successfully based the selection of remedial procedures upon this system of errors, a number of improvements could be made. It is apparent that his types would permit the classification of an error in several categories, *i.e.*, his samples of unfinished words are certainly omissions of letters. They might also be called phonetic misspellings, inattention to latter part of word, or omissions to shorten word. Subjective classifications are freak errors, words not recognized, similar graphic responses, and repetition of dominant elements. Wells fails to make distinctions between failure to double and incorrect doubling or between phonetic and non-phonetic additions and omissions.

In 1935 Watson[40] reported the results of a series of experiments in the psychology and pedagogy of spelling. In one of these she proposes a system of diagnosis and remedial work for aiding those disabled in spelling. Her diagnosis includes an analysis of spelling errors. The errors or, rather, the spelling habits that she would note are:

I. Attending to syllables—the habit of noticing all the syllables of a word by listening, visually observing, pronouncing and writing.
   A. Rhythmic pattern of a word—the habit of separating into syllables with short pauses and accents indicating syllabic rhythms.

B. Differentiation of syllables—giving correct emphasis to the syllables according to their relative importance within the word.

C. Phonic quality of every syllable—the habit of perceiving all the phonic elements of a syllable.

(To note these spelling habits, the examiner would take account of such errors as omissions of syllables or letters, partial reversals or transpositions, and phonetic substitutions for consonant and vowel sounds.)

II. Establishing specific new or unique spelling reactions—the habit of drill on new or unique words with attention to syllables.

III. Transferring learned bonds according to valid system of analogies where common elements occur—making use of transfer of knowledge of phonics; spelling by correct analogy.

A. Knowing what are possible ways of spelling a given sound-unit.

B. Associating specific word-parts with their proper sound-unit categories.

1. ei for ie for long e sound.
2. Change final y to i before a suffix?
3. os or oes in plural.
4. Drop final e when suffix is added?
5. c and g soft before e, i and y.
6. Double final consonant before a suffix?
7. ful as a suffix.
8. ness as a suffix.
9. some as a suffix.
10. ment as a suffix.
11. ed as a suffix (especially when pronounced as t).

IV. Forming letters legibly in handwriting.

V. Coördination and motor control.

1. Avoiding lapses or slips of mind or pen.
2. Overcoming constant errors of order, substitution, insertions, or omissions.

Watson's second topic cannot be observed by the method of tabulating spelling errors. Her third topic, the child's use of transfer of his knowledge of phonics may be applied by noting spelling errors. She indicates that the examiner would look for error of phonetic spelling resulting in combinations that are not used in English to represent the sound substituted for, *i.e.*, "wild" or "impossible" spellings. One would also note whether the pupil is employing a spelling possible for the sound but incorrect for the word in question. From a pragmatic viewpoint such an analysis is almost impossible. It has been estimated that our vowels have a total of one hundred four possible sounds. Similarly, common two-letter combinations have an average of five to six different sounds or pronunciations.* The labor

* Spache, George: "A Phonics Manual for Primary and Remedial Teachers." *Elementary English Review*, XVI, April and May, 1939, pp. 147–150, 191–198.

necessary to determine whether the child's spelling resembles any sound-unit in English is unwarranted, in the writer's opinion.

Section B of Watson's third topic is a partial list of spelling conventions or rules. As previously observed, this is not a study of spelling errors in the sense employed here. In her fourth and fifth categories, Watson would note, respectively, poor letter formation, letter eccentricities and errors of partial reversals, substitutions, additions, and omissions.

> The basic psychological assumption underlying the scheme is that, regardless of what may have been the original cause of a special spelling disability, its nature in effective reality is a lack of specific detailed habits and organizations of habits concerned in the correct writing of English words ([40], p. 45).

In Watson's own words, this is the *raison d'etre* of her classification scheme. The technique is in effect an analysis of the habits of the student and his knowledge of a number of spelling rules. As such it is a distinct improvement upon preceding studies. However, in the hands of other users than its author the scheme may not prove practical for the tabulation of spelling errors because of the large element of subjectivity. It is obvious that several of the types of habits may appear in similar mistakes, *i.e.*, partial reversals may be lack of attending to syllables or lapses of mind or pen. Like other writers employing such analyses Watson does not make clear how one is to distinguish between accidental and significant errors.

The dependence on a few specific spelling rules is as yet unwarranted. Research has not yet definitely determined the relative values of all the rules offered in orthography. There may be a number of rules violated by certain errors that many writers would claim to be significant. One also doubts that there is equal opportunity for violation of the various rules in any of the spelling tests used to collect spelling errors.

Psychologically, Watson's analysis is the second step in diagnosis. The necessity for a detailed analysis of errors before analytic grouping and interpretation has already been emphasized. In this, the writer is following the example of the systems of diagnosis and remedial work employed by Marion Monroe,[46] William Gray[47] and Durrell.[12] These make an attack upon the analogous problem of reading disability from an initial tabulation of the specific errors of the pupil. The efficiency of such an approach has been amply demonstrated by the success and widespread use of these systems.

In a study of the spelling of adults, Schonell[33] noted only errors of omission, transposition, substitution, and addition. None of these were differentiated in terms of their constituent types.

The Gates-Russell Spelling Diagnosis Tests[16] suggest that the teacher note additions, insertions, omissions, substitutions, transpositions, and phonetic errors. The list is identical with that used by Russell in his, *Characteristics of Good and Poor Spellers*.[31] These authors also give Mendenhall's[26] listing as an alternate "more exact listing." It is their opinion that, "A list as specific as this (Mendenhall's) enables the teacher to analyze just what types of error a pupil is making. On the basis of such an analysis remedial work can be started on the definite points where it is most needed" ([16], p. 9).

The present writer cannot concur in the belief that either of these lists is "exact" or "specific." In his opinion, they are characterized by superficiality, in the case of the Gates-Russell system, and by redundancy and inaccuracy, in the case of the Mendenhall list.

In his *Analysis of Reading Difficulty*, Durrell[12] suggests analysis of the spelling errors of the subject in terms of the following: Omits or adds sounds or syllables, incorrect phonetic spelling, correct phonetic spelling but incorrect on non-phonetic elements, slow handwriting. He also suggest classifying spelling errors according to these types, which he also employs in analysis of faulty pronunciation:

Vowel errors—phonetic, non-phonetic.
Consonant errors—phonetic, non-phonetic.
Reversals—b-d, p-q; in sequence.
Additions of sounds—beginning, middle or end of word.
Omissions of sounds—beginning, middle or end of word.
Substitution of whole word—similar form, similar idea.

Durrell's classification might be improved by distinguishing the type of error in vowels and consonants in terms of omissions, addition, and substitutions; by noting phonetic and non-phonetic reversals, confusion of homonyms and errors due to pupil mispronunciation.

To give some idea of the confusion existing in the classification of spelling errors, the various causes to which errors have been attributed may be summarized:

Omissions—haste in writing—Vigilia
regressive or associo-suggestive assimilation—F. L. Wells
phonetic spelling of irregular words—Rusk
because of obscure or elided vowels and consonants—Cook and O'Shea

similar motor response—Hollingworth
sense of having written element—Lester
because last of one word and first of next have same or similar sound—Hollingworth

Additions—haste in writing—Vigilia
sensori-motor attraction—Cornman
repetition of dominant elements—Hollingworth
perseveration—Hollingworth
kinesthetic inertia—Speer
tendency to anticipate—Book and Harter

Transpositions—anticipation, regression—F. L. Wells
faulty visual analysis—Rusk

Substitutions—ideo-motor attraction—Cornman
regression, incorrect analogy—F. L. Wells
confusing similar graphic responses—Hollingworth
common visual, kinesthetic or phonetic elements—Hollingworth
forgetting word—Book and Harter

Handwriting—very slow rate—Floyd, Durrell
dwarfing letter, letter formation—F. L. Wells
confusing letters above and below line—Cook and O'Shea
poor penmanship—Speer
haste in writing—Vigilia
letter eccentricities—Watson

Miscellaneous general causes (used to explain any or all of the preceding).
mispronunciation, foreign or colloquial pronunciation.
faulty visual analysis or image, confused imagery.
lack of knowledge or familiarity with word.
carelessness, random performance, forgetting word.
lack of attention, careless teaching.
idiosyncrasies.
misunderstanding, mishearing, failure to hear or perceive.
defective hearing, vision or speech; fatigue.
tendency to form familiar word or word parts.
lack of knowledge or application of phonics; the hindrance of phonics; lack of knowledge of sounds.
spelling by phonetic analogy.
lack of knowledge of certain spelling rules or conventions.
lapses of various kinds.

In the writer's opinion an adequate classification should fulfill the following requirements:

(1) It should include all types of errors in significance in understanding the pupil's spelling.

(2) The types of errors should include all characteristic tendencies of children in their attempts to spell, *i.e.*, phonetic spelling, etc.

(3) The types of errors should be clearly delimited to prevent overlapping and to avoid confusing the examiner.

(4) The rules for the classification of errors should be simple enough for one to apply the technique accurately after a small amount of practice.

(5) The rules should be sufficiently objective so as to leave little to the judgment or opinion of the user as to the correct classification of an error.

(6) The classification should not be an attempt to explain cause but an objective method of noting the detailed characteristics of the errors.

(7) The classification should permit grouping of types of errors after the tabulation to aid the examiner in understanding the spelling habits and knowledges of the pupil.

An adequate system of classification should meet the following statistical requirements:

(1) The validity of the technique as an accurate tabulation of the pupil's errors should be established.

(2) The reliability of the system as a whole, and of the major and minor types of error should be established.

(3) The significance of the frequency of errors of a particular type should be established.

(4) The number of misspelled words necessary to secure an adequate, reliable sampling of the pupil's spelling should be determined.

(5) The basis of the collection of errors should be a reliable standardized spelling scale with norms established by widespread testing, rather than a random collection of misspellings from the child's notes, compositions, letters, etc. This will facilitate the determination of the grade level of achievement and the quantitative extent of the disability simultaneously with the analysis of the spelling errors.

(6) The spelling analysis should be based upon at least one hundred words from a fifty per cent level of difficulty, *i.e.*, a level at which approximately fifty per cent might be expected to be correct according to the grade level of the pupil.*

The writer offers the following system and rules for classifying spelling errors in the belief that it fulfills the initial requirements. The information necessary to fulfill the statistical requirements will be presented in a successive article. The writer hopes also to complete the diagnostic technique by describing the testing of such visual, auditory, language and motor abilities as research has shown to be clearly fundamental to spelling ability. The complete diagnosis will also take cognizance of such factors as poly-lingualism, intelligence, developmental data, medical data and school history.

In order to acknowledge the derivation of the writer's system, and

---

* Both Otis[28] and Starch[35] have indicated that a list of twenty words gives a very poor measure of ability in spelling. Otis would use one hundred words, while Starch feels two hundred are necessary. The most reliable measure is obtained by using words for which there is an average correct spelling of fifty per cent.[28] The Iowa Spelling Scales and the Buckingham Revision of the Ayres Spelling Scale would seem particularly suited for such testing.

to aid the reader in summarizing the preceding research, the types of errors are attributed to the earliest writer in whose work they appeared or by whom they were most clearly phrased. The originator of the type is listed immediately after the name of the error. In some instances a second author is cited because of significant addition to the definition of the error type.

RULES FOR THE CLASSIFICATION OF MISSPELLED WORDS

General.

1. A maximum of three errors in a single word will be tabulated.
2. No specific error will be entered more than once.

Omissions.

An omission is one in which a letter or letters of the test word are omitted in the child's attempt at spelling, without any attempt at substitution.

The omission of two or more successive letters is considered as a single error and tabulated under the omission of a syllable.

Single letter.

silent letter—(Foster)—the omission of a consonant or vowel that is silent in the ordinary pronunciation of the word. This includes silent final e and the silent or subordinate vowel of a diphthong. As wether-weather, reman-remain, fin-fine

sounded letter—(Gill)—the omission of a letter that is sounded in the ordinary pronunciation of the word. This includes the omission of the major or sounded vowel of a diphthong. As requst-request, plasure-pleasure, personl-personal, juge-judge.

Doubled letter—(Cornman)—the omission of one of a pair of successive, identical letters. As suden-sudden, adress-address, sed-seed.

Syllable—(Douse)—the omission of two or more letters forming a syllable. As (en)joy, preced(ed), import(an)ce.

Additions or Repetitions.

An addition is one in which a letter or letters are added. The correct letters of the test word are present in the portion of the word in which the error of addition occurs. A repetition of a letter or letters of the test word is considered an addition.

Single letter.

doubling—(Cornman)—the repetition of a single letter. As untill-until, frriend-friend, deegree-degree

non-doubling.

phonetic—the addition of a letter that might appear in a phonetic spelling of the word. This includes using a diphthong for a vowel, as ea for short e, ai for a, ie for long i, etc. As shead-shed, compainy-company, retiere-retire. Includes also the addition of a consonant, as ck for c. As eleckt-elect, pleashure-pleasure.

non-phonetic—the addition of a letter that would not appear in a phonetic spelling of the word. As untile-until, empirer-empire.

Syllable—(Foster, Cornman)—the addition of two or more letters. As gathered-gather, accidentir-accident. Also the repetition of a syllable of the test word, as unusual-usual.

Transpositions or Partial Reversals.

A transposition or partial reversal is one in which the correct sequence of the letters of the test word is disturbed. It usually takes the form of reversing the sequence of two letters. It may also take the form of interchanging the letters on either side of a central letter, as was-saw. Includes the transposition of syllables.

Phonetic—transposition in an attempt to spell the word phonetically or the reversing of such common phonetic alternatives as el-le, ei-ie. As cride-cried, nickle-nickel, empier-empire.

Non-phonetic—transposition not resulting in phonetic spelling. As baord-board, appaer-appear.

Phonetic Substitutions.

A phonetic substitution is one in which a letter or letters similar in sound to those of the test word are substituted. The letters of the test word for which the substitution is offered are, of course, omitted.

Vowel sound—(Gill)—substitution of a vowel or vowel and consonant for a vowel of the test word. As prisin-prison, injoy-enjoy.

Consonant sound—(Hollingworth, Brandenburg)—substitution of a consonant or vowel and consonant for a consonant of the test word. As prizon-prison, cecond-second, vakation-vacation.

Diphthong sound—(Murray)—substitution of a vowel (not present in the diphthong of the test word) or of a vowel and consonant or diphthong for the diphthong of the test word. As Ogust-August, certonly-certainly, unfained-unfeigned.

Syllable—(Brandenburg)—substitution of a similarly sounding syllable for a syllable of the test word. As purchest-purchased, financhel-financial. Includes also the substitution of a single letter for a syllable. As naborhood-neighborhood, stopt-stopped.

Entire word—(Rusk)—the substitution of an actual word grossly similar in sound. As weary-very, colonial-colonel.

Non-phonetic Substitutions.

A non-phonetic substitution is one in which a letter or letters dissimilar in sound are substituted in the test word. The letters of the test word for which the substitute is offered are, of course, omitted.

Vowel—(Cornman)—substitution of a vowel or vowel and consonant for the vowel of the test word. As stition-station, strick-struck.

Consonant—(Cornman)—substitution of a consonant or vowel and consonant for a consonant of the test word. As wathcing-washing, inportance-importance.

Diphthong—substitution of a vowel or consonant or both or of a diphthong for a diphthong of the test word. As unfrined-unfeigned.

Syllable—(Douse, F. L. Wells)—substitution of a syllable or a vowel or consonant for a syllable of the test word. As difficould-difficult.

Entire word—(Rusk, Masters)—substitution of an actual, irrelevant word. As wee-clean, born-lord, urge-arrange.

Confusing Homonyms—(Stead).

An error in which the pupil substitutes a word identical in sound for the test word. As two-too, slay-sleigh, cause-course.

Carelessness—(Foster).

A careless error is one in which a word known by the pupil is misspelled. It can only be determined by having the child spell test words orally. If again misspelled, tabulate errors.

Handwriting—(F. L. Wells).

Error obviously due to motor incoördination or poor letter formation. Include illegible letters.

Mispronunciation—(Floyd).

Check pronunciation of all words tested by having pupil read list of test words. Use typewritten or printed list rather than pupil's list. Note similarities between mispronunciation and misspelling.

Words Refused or Not Completed—(Floyd).

Entire test word is omitted or only partly completed. As libt-liberty, di-decided, sud-sudden.

Unrecognizable—(Bugbee, Umberhine).

Combinations not forming a real or misspelled word, unrecognizable as words. Include words with four or more errors. As surgar-struggle, instard-instructor.

It is apparent that the greater part of the writer's proposed system of classification is derived from numerous sources. Types of errors not previously used include omission of the silent or subordinate vowel in a diphthong; omission of the sounded or major vowel in a diphthong; additions of a single letter representing an attempt at phonetic spelling; non-phonetic single additions; phonetic transpositions distinguished from non-phonetic and non-phonetic substitutions for a diphthong.

The present writer hopes that this work represents a step toward greater objectivity of analysis of spelling errors, first, by an elimination of subjective types of errors; second, by a careful distinction between errors and causes; and, third, by a careful definition of each error type. He welcomes critical analysis of the proposed system of classification of spelling errors.

BIBLIOGRAPHY

1. Book, W. F. and Harter, R. S.: "Mistakes Which Pupils Make in Spelling." *Journal Educational Research*, Vol. XIX, February, 1929, pp. 106–118.
2. Brandenburg, G. C.: "The Spelling Ability of University Students." *School and Society*, Vol. IX, January 5, 1918, pp. 26–29.
3. Breed, Fred. S.: "Spelling." *Journal of Educational Research*, Vol. V, February, 1935, pp. 83–88.
4. Bugbee, Ethel Schurman: *The Influence of Phonics upon Spelling.* Master's thesis, University of Iowa, 1938.
5. Burt, Cyril: *Mental and Scholastic Tests.* London: P. S. King & Son, Ltd., 1927, pp. 432.
6. Carle, Rose Mary: *An Analytical and Comparative Study of the Spelling Deficiencies of Freshmen and Upper-Classmen in the University of Kansas.* Master's thesis, University of Kansas, 1928, pp. 115.
7. Carroll, Herbert A.: *Generalization of Bright and Dull Children.* Teachers College, Columbia University Contributions to Education, No. 439, 1930.
8. Cook, W. A. and O'Shea, M. V.: *The Child and His Spelling.* Indianapolis: Bobbs Merrill, 1914, pp. 254.
9. Cornman, Oliver P.: *Spelling in the Elementary School.* New York: Ginn & Co., 1902, pp. 98.
10. Davis, Georgia: "Remedial Work in Spelling." *Elementary School Journal*, Vol. XXVII, April, 1927, pp. 615–626.
11. Douse, T. L.: "On Some Minor Psychological Inferences: A Study of Misspellings and Related Mistakes." *Mind-New Series*, Vol. IX, 1900, pp. 85–93.
12. Durrell, Donald D.: *Durrell Analysis of Reading Difficulty.* Yonkers: World Book, 1937, pp. 24.

13. Floyd, Hazel: *Cases of Spelling Disability: Their Diagnosis and Treatment.* Master's thesis, University of Chicago, 1925, pp. 221.
14. Foran, T. G. and Rock, R. T.: "An Annotated Bibliography of Studies Relating to Spelling: Supplement No. 1." Catholic University, *Educational Research Bulletin*, Vol. v, No. 1, 1930, pp. 24.
15. Foster, Wm. Trufant: "The Spelling of College Students." *Journal Educational Psychology*, Vol. II, 1911, pp. 211–216.
16. Gates, A. I. and Russell, David H.: *Diagnostic and Remedial Spelling Manual.* New York: Teachers College, Columbia University, 1937.
17. Gill, E. J.: "The Teaching of Spelling." *Journal Experimental Pedagogy*, Vol. I, 1912, pp. 310–319.
18. Hill, Ina H.: "Diagnosis of Spelling Difficulties." *Elementary English Review*, Vol. I, November, 1924, pp. 225–230.
19. Hollingworth, Leta S.: *The Psychology of Special Disability in Spelling.* Teachers College, Columbia University, Contributions to Education, No. 88, 1918, pp. 105.
20. Irmina, Sister M., Visitation, Sister M. and Gabriel, Sister M.: "An Annotated Bibliography of Studies Relating to Spelling." Catholic University, *Educational Research Bulletin*, Vol. III, No. 1, 1928, pp. 56.
21. Kline, Linus W.: "A Study in the Psychology of Spelling." *Journal Educational Psychology*, Vol. III, September, 1921, pp. 381–400.
22. Leland, Bernice: *Your Pupils and Mine: Case Studies in Reading, Arithmetic and Spelling.* Detroit Teachers College, 1929, pp. 69.
23. Lester, John A.: "A Study of High School Spelling Material." *Journal Educational Psychology*, Vol. XIII, 1922, pp. 65–75, 152–159.
24. Lester, John A.: *A Spelling Review.* Doylestown, Pa.: Lester Publishing Co. 1917, pp. 110.
25. Masters, Harry V.: "A Study of Spelling Errors." *University of Iowa Studies in Education*, Vol. IV, No. 4, September, 1927, pp. 80.
26. Mendenhall, James A.: *An Analysis of Spelling Errors.* Lincoln School, Teachers College. New York: Teachers College, Columbia University, 1930.
27. Murray, Elsie: "The Spelling Ability of College Students." *Journal Educational Psychology*, Vol. X, October, 1919, pp. 357–376.
28. Otis, Arthur S.: "The Reliability of Spelling Scales." *School and Society*, Vol. IV, 1916, pp. 676–683, 716–722, 750–756, 793–796.
29. Randall, C. F.: "The Teaching of Spelling." *Popular Educator*, Vol. XLI 1924, pp. 332–335.
30. Rusk, R. R.: "An Analysis of the Spelling Errors of Adults." *Journal Experimental Pedagogy*, Vol. II, 1913, pp. 119–122.
31. Russell, David H.: *Characteristics of Good and Poor Spellers: A Diagnostic Study.* Teachers College, Columbia University, Contributions to Education, No. 727, 1937, pp. 103.
32. Sartorius, Ina Craig: *Generalization in Spelling.* Teachers College, Columbia University, Contributions to Education No. 472, 1931, pp. 65.
33. Schonell, Fred. J.: "Ability and Disability in Spelling Among Educated Adults." *British Journal Educational Psychology*, Vol. VI, 1936, pp. 123–146.

34. Speer, Dorothy: "An Analysis of the Sources of Errors in One Hundred Four Words Misspelled by Primary-School Children in the City of Baltimore." *Elementary School Journal*, Vol. XXVIII, January, 1928, pp. 362–372.
35. Starch, Daniel: "The Measurement of Efficiency in Spelling and the Overlapping of Grades in Combined Measurement of Reading, Writing and Spelling." *Journal Educational Psychology*, Vol. VI, March, 1915, pp. 167–186.
36. Stead, Harry G.: "Spelling Errors in Children." *Journal Experimental Pedagogy*, Vol. II, 1914, pp. 362–364.
37. Sudweeks, Joseph: "Practical Helps in Teaching Spelling: Summary of Helpful Principles and Methods." *Journal Educational Research*, Vol. XVI, September, 1927, pp. 106–118.
38. Umberhine, Ethel Marie: *Comparison of Errors of Good and Poor Spellers.* Master's thesis, University of New Mexico, 1933, pp. 221.
39. Vigilia, Sister M.: "Some Values of Diagnosis in Spelling." *Elementary English Review*, Vol. III, February, 1926, pp. 39–41.
40. Watson, Alice E.: *Experimental Studies in the Psychology and Pedagogy of Spelling.* Teachers College, Columbia University Contributions to Education, No. 638, 1935, pp. 144.
41. Welch, Ralph W.: *Spelling Errors of Certain California High School Seniors.* Master's thesis, University Southern California, 1928, pp. 83.
42. Wells, F. L.: "Linguistic Lapses." *Archives of Philosophy, Psychology and Scientific Methods*, No. 6, 1906.
43. Wells, George N.: *The Effect of Specific Remedial Procedures in Spelling.* Master's thesis, Department of Education, University of Chicago, 1933, pp. 190.
44. Williamson, Edmund G.: "Mental Abilities Related to Learning to Spell." *Psychological Bulletin*, Vol. XXX, December, 1935, pp. 743–751.
45. Wycoff, Glenn L.: *An Analytical Study of the Spelling Deficiencies of High-School Seniors in Twenty-six Representative High Schools in Kansas and Oklahama During the School Year* 1925–26. Master's thesis, University of Kansas, 1928, pp. 58.
46. Monroe, Marion: *Children Who Cannot Read.* Chicago: University of Chicago Press, 1932, pp. 205.
47. Gray, Wm. S.: *Standardized Oral Reading Check Tests.* Bloomington, Illinois: Public School Publishing Co.

FORM OF STATEMENT IN PERSONALITY MEASUREMENT*

EDWARD A. RUNDQUIST

Psychological Laboratory, Cincinnati Public Schools

It is the purpose of this paper to call attention again to a factor which is a source of marked confusion in interpreting the results of pencil-and-paper personality measurement, and to point out the potential value of studying the characteristics of the responses to personality or attitude items, a phase of study which has received little attention. There are many such factors which have not been given sufficient study: Influence of the position of an item in a series, the number of alternative responses allowed for each item, the general technique of statement, and the form of statement. This last factor is our concern in the present study.

The forms of statement considered are those termed acceptable and unacceptable by Rundquist and Sletto (7). In that study, acceptable items were defined as those, agreement with which indicates a favorable position with respect to the trait measured; unacceptable items as those, agreement with which indicates an unfavorable position with respect to the trait under consideration. In other words, acceptable items state the favorable side of the case—"Home is the best place in the world"; while unacceptable items state the unfavorable side of the issue—"Home is the worst place in the world."

Such items have been termed positive and negative by other investigators, a terminology abandoned in the above study because the terms were not considered descriptive (7, 243–245). Maller (5) used both forms in his *Character Sketches* with the view of detecting consistency of response. Randolph Smith (8), however, was the first to investigate planfully the influence of these forms of statement on the response of the individual. He attempted to control ideational content of the item but, for the most part, confined his analysis to differences in response when the two types of statement were administered in separate forms.

Smith showed clearly that the influence of form of statement was less when the two types of statement were arranged at random in a single form. At the outset of Rundquist and Sletto's study, form of

---

* Presented, in part, at the September 1937 meetings of the American Psychological Association.

statement was not considered a problem where the two kinds of statement were administered in a single form, and hence no attempt was made to control the ideational content of the items. Even when the two types were administered in a single form, the importance of the problem became evident, and the findings in this respect are among the more significant that emerged from the study.

It seemed desirable, therefore, to make a critical test of the influence of form of statement, when both forms were included in a single scale, by controlling ideational content of the item.

One hundred personality items were administered to two successive groups of eighth-grade pupils of a large school—each year's class numbering about three hundred fifty. Forty of the items concerned family or home adjustment; twenty of these were acceptably stated, while the remaining twenty were identical in content but unacceptably stated.

Five response positions were provided for each item. The following two examples will illustrate the general set-up of the scales:

HOME IS THE BEST PLACE IN THE WORLD.

Strongly Agree[5] Agree[4] Undecided[3] Disagree[2] Strongly Disagree[1]

HOME IS THE WORST PLACE IN THE WORLD

Strongly Agree[1] Agree[2] Undecided[3] Disagree[4] Strongly Disagree[5]

The items were scored according to the arbitrary weights assigned to each response position, a technique which the work of Likert (3) and Rundquist and Sletto (7) justifies. The low weight was assigned to the unfavorable end of the scale of responses.

The twenty acceptable family adjustment items together with their unacceptably stated opposites are given below:

ACCEPTABLE ITEMS	UNACCEPTABLE ITEMS
It is easy to get along at home.	It is hard to get along at home.
It is easy to be cheerful at home.	It is hard to be cheerful at home.
Parents treat you like a grown-up.	Parents treat you like a baby.
Parents do *not* expect too much.	Parents expect too much.
It is easy to behave at home.	It is hard to behave at home.
Parents remember how it feels to be young.	Parents forget how it feels to be young.
Parents let you choose your own friends.	Parents do *not* let you choose your own friends.
One hardly ever becomes nervous at home.	One becomes nervous at home.

ACCEPTABLE ITEMS	UNACCEPTABLE ITEMS
Home is the best place in the world.	Home is the worst place in the world.
It will be fun to have a boy or girl of your own.	It will *not* be fun to have a boy or girl of your own.
It is fun to help around the house.	It is no fun to help around the house.
It is easy to behave so your family is satisfied.	It is hard to behave so your family is satisfied.
It is easier to have a good time at home than anywhere else.	It is harder to have a good time at home than anywhere else.
It is easy to please your family.	It is hard to please your family.
It is easy to tell your troubles to your parents.	It is hard to tell your troubles to your parents.
Parents give good advice.	Parents give poor advice.
Brothers and sisters are nice to have around.	Brothers and sisters are nearly always in the way.
You should tell your parents everything you do.	You should *not* tell your parents everything you do.
Parents understand you better than anyone else.	Parents do not understand you as well as other people do.
People in the family are *not* too curious about what you are doing.	People in the family are too curious about what you are doing.

Inspection of these pairs will show that the ideational content of the items is as rigorously controlled as is possible within the limits of the English language. This control was achieved by using opposites: Best, worst; easy, hard; or, where the statement did not lend itself to such treatment, by inserting a negative into one of the paired statements.

The remaining sixty items administered were designed to measure inferiority, attitude toward school, and attitude toward the future or morale. There were twenty items in each of these scales, ten of each form of statement but not opposite in ideational content. The two series of family items are identical in content. Therefore, a comparison of the two series in their relationships to the remaining series of items will yield a measure of the influence of form of statement.

These relationships are presented in Table I. The numbers in parentheses are column numbers. The letter *U* stands for unacceptable, the letter *A* for acceptable. In column 1 are the correlations of the *U* or unacceptable Family items with the *A* or acceptable items of the remaining scales; in column 2 are the correlations of the *U* or unacceptable Family items with the *U* or unacceptable items of the other scales; and so on. The results are highly pleasing in their con-

sistency. Table I shows no deviation from the trends to be discussed in terms of the average values of the correlation coefficients in each column.

TABLE I.—RELATIONSHIP OF TWO SERIES OF FAMILY ADJUSTMENT ITEMS, DIFFERING SOLELY IN FORM OF STATEMENT, TO OTHER PERSONALITY MEASURES
(Subjects were eighth-grade children)

	1935–1936 boys (N = 172)				1935–1936 girls (N = 186)			
	*U** family items with		*A** family items with		*U* family items with		*A* family items with	
	*A**	*U*†	*A**	*U*†	*A*	*U*	*A*	*U*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Morale items	.32	.60	.57	.39	.34	.73	.50	.44
Inferiority items	.27	.55	.48	.40	.28	.66	.48	.42
Education items	.34	.47	.47	.41	.18	.67	.36	.49
Mean	.31	.54	.51	.40	.27	.69	.45	.45
	1936–1937 boys (N = 177)				1936–1937 girls (N = 164)			
	U family items with		*A* family items with		*U* family items with		*A* family items with	
	A	*U*	*A*	*U*	*A*	*U*	*A*	*U*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Morale items	.29	.59	.36	.39	.14	.63	.44	.44
Inferiority items	.22	.51	.40	.31	.16	.62	.43	.42
Education items	.14	.34	.26	.21	.07	.59	.28	.34
Mean	.22	.48	.34	.30	.12	.61	.38	.40

* Acceptable.
† Unacceptable.

Column 2 shows the average of the correlations between the Family unacceptable items and the unacceptable Morale, Inferiority, and Education items; column 4, the Family acceptable items with these same unacceptable items. In other words, the only factor which is different in the two sets of correlations is the form of statement of the

Family items. For boys of the 1935–1936 group, the average is .54 when unacceptable Family items are correlated against other unacceptable items, and .40 when acceptable Family items are correlated against the same unacceptable items of the other scales. Corresponding values for the 1936–1937 data are .48 and .30. For girls the values in columns 6 and 8 give the same comparison. These are .69 against .45 for the 1935–1936 data and .61 against .40 for the 1936–1937 data. Unacceptable items of the Family scale clearly correlate higher with unacceptable items of the other scales than do acceptable Family items of identical ideational content.

Columns 1 and 3 for boys and 5 and 7 for girls likewise yield comparisons in which form of statement is the only variable. It may be seen that the acceptable Family items correlate higher with acceptable items of the remaining scales than do unacceptable Family items—the values being .51 (column 3) against .31 (column 1) for the 1935–1936 boys; .34 against .22 for the 1936–1937 boys; .45 (column 7) against .27 (column 5) for the 1935–1936 girls; and .38 against .12 for the 1936–1937 girls.

In the comparisons of Table I the variables with which the two series of Family items were correlated have a clearly different ideational content. If these variables were likewise composed of Family items differing only in specific detail, would form of statement still be a factor? Or would the general similarity of ideational factor overcome its influence? The forty items of the Family series can be divided into four groups to provide an answer to this question. The twenty acceptable items were divided into two series to be termed FA_{I} and FA_{II}; the twenty unacceptable items were likewise divided into two parts, FU_{I} and FU_{II}. The specific ideational content of FA_{I} and FU_{I} is identical, as is that of FA_{II} and FU_{II}. FA_{I} and FU_{I} differ solely in form of statement, as do FA_{II} and FU_{II}. There are thus two series differing solely in form of statement, which may be compared in their relationships with other variables. Table II is arranged to give these comparisons. In the upper half of the table FU_{I} and FA_{I} are compared in their relationships to FU_{II} and FA_{II}; in the lower half, FU_{II} and FA_{II} are compared in their relationships to FA_{I} and FU_{I}. The lower half of the table is simply a rearrangement of the upper half to facilitate the comparisons. Comparison of columns 1 and 3, and of 2 and 4, brings out the same points noted from Table I; namely, that of two series of items differing solely in form of statement, each will correlate higher with a series of like form than it will with

one of opposite form of statement. The most critical comparisons we can make clearly demonstrate the influence of form of statement when both forms are arranged at random in a single scale.

Having established this point, it is of interest to inquire whether other phenomena associated with form of statement are consistent with the findings of Smith (8) and Rundquist and Sletto (7). It was

TABLE II.—INTERCORRELATIONS OF FOUR SECTIONS OF THE FAMILY ITEMS
(See text for explanation)

	N	FU_I with		FA_I with	
		FA_{II} (1)	FU_{II} (2)	FA_{II} (3)	FU_{II} (4)
1935–1936 boys	172	.48	.58	.61	.51
1935–1936 girls	186	.49	.76	.60	.52
1936–1937 boys	177	.61	.56	.70	.45
1936–1937 girls	164	.58	.73	.67	.55
Mean	...	.54	.66	.64	.51
		FU_{II} with		FA_{II} with	
		FA_I (1)	FU_I (2)	FA_I (3)	FU_I (4)
1935–1936 boys	172	.51	.58	.61	.48
1935–1936 girls	186	.52	.76	.60	.49
1936–1937 boys	177	.45	.56	.70	.61
1936–1937 girls	164	.55	.73	.67	.58
Mean	...	.51	.66	.64	.54

noted in the latter study that unacceptable items tend to intercorrelate higher than do acceptable. Tables I and II show this tendency for girls but not so consistently for boys. In this respect the present study is not completely consistent with that of Rundquist and Sletto (7), who found the tendency for *U* items to intercorrelate higher than *A* items very pronounced for men. The difference in age of the groups dealt with in the two studies may be the explanation of this discrepancy. In both studies the evidence is very clear and consistent, so it is improbable that either study errs in its description of the facts.

In the previous studies (7, 8) a clear tendency is noted for mean scores of unacceptable items, relative to acceptable item means, to tend more toward the favorable end of the scale, and for standard deviations of unacceptable item series to be larger. Table III con-

TABLE III.—MEANS AND STANDARD DEVIATIONS FOR ACCEPTABLE (*A*) AND UNACCEPTACLE (*U*) ITEMS OF THE FOUR SCALES

1935–1936

	Boys (N = 172)		Girls (N = 186)	
	Mean*	σ	Mean	σ
Family *A* items	73.66	9.03	76.30	9.04
Family *U* items	76.53	9.70	79.66	11.20
Education *A* items	31.04	6.47	34.11	4.32
Education *U* items	32.94	6.38	36.82	5.79
Inferiority *A* items	33.37	4.57	32.98	4.73
Inferiority *U* items	30.81	5.68	32.71	5.51
Morale *A* items	34.99	4.99	34.69	4.33
Morale *U* items	37.24	5.42	37.69	5.71

1936–1937

	Boys (N = 177)		Girls (N = 164)	
	Mean	σ	Mean	σ
Family *A* items	74.84	10.57	77.58	8.87
Family *U* items	78.06	10.40	80.60	11.43
Education *A* items	32.51	5.99	35.50	4.52
Education *U* items	34.19	6.37	36.43	5.78
Inferiority *A* items	34.23	4.15	33.54	4.54
Inferiority *U* items	32.48	5.00	33.38	5.79
Morale *A* items	36.05	4.43	35.63	4.63
Morale *U* items	38.93	5.34	38.45	5.40

* High scores represent the favorable end of each scale.

firms these trends for the present data. Discussion of the few exceptions seems unprofitable at this time.

It should be noted, however, that, once a discrepancy can be determined to be a real one, knowledge of the meaning to be attached to the findings concerning form of statement will clarify interpretation. This may be illustrated by the tendency of *A* Family items to correlate

with *U* items of other scales to a greater extent than do Family *U* items with *A* items of other scales. (See Table I.) For boys, compare columns 4 and 1. The average values are .40 and .31 for the 1935–1936 data; .30 and .22 for the 1936–1937 data. For girls, compare columns 8 and 5. The average values are .45 to .27 for the 1935–1936 data, and .40 to .12 for the 1936–1937 data. The content of the item or, perhaps, the adjustment area being tapped, is apparently a factor in the degree to which form of statement operates. As more becomes known concerning the basic reasons for the differential response to the two types of item, this kind of finding may serve as a clue to the nature of the trait being tapped by any particular scale. It may mean, in this instance, that the Family items are tapping a trait or traits of more highly charged emotional significance to the individual. In the present state of our knowledge such an inference is not completely justified. It is made merely to call attention to the type of problem that is met with in this kind of analysis and to indicate one of the ways in which knowledge concerning form of statement may assist in clarifying the field of personality measurement.

Taken together, the work of Smith (8), Rundquist and Sletto (7), and the present study provide indisputable evidence of the influence of form of statement, whether the two forms are administered in a single scale or not. The interpretations to be put on these findings are, however, by no means clear. One finding seems established, however; unacceptable items possess greater validity in the sense of discriminating better between criterion groups.

Rundquist and Sletto (7) found that men receiving public relief scored significantly less favorably on thirty-one items of the Minnesota Survey of Opinions. Of these thirty-one items, twenty-four were unacceptably stated. Similar evidence is found in studies by Hall (2) and Pintner, Maller, and Forlano (6).* Hall published the critical ratios for fifty-seven item differences between a group of three hundred sixty unemployed and three hundred employed engineers. Twenty-six of these items were acceptably stated; thirty-one, unacceptably stated. The average critical ratio for the acceptable items was 1.50; the average for the unacceptable, 4.01.

---

* The computations presented were made by the present writer from the data presented by these authors. It is sometimes difficult to classify an item as acceptacle or unacceptable in the absence of the scoring key. The writer's classification statements by Pintner's *et al.* may be slightly in error. In general, however, he is certain of the correctness of his classification.

Pintner, Maller, and Forlano give the critical ratios for two hundred item differences* between a group of problem and a group of normal children. The average critical ratio for the ninety-two acceptable items was 2.50; the one hundred eight unacceptable items, 3.09.

While it is clear from the evidence presented that unacceptable items tend to be the more valid, certain evidence implies that the problem is by no means this simple. Nor does it follow that acceptable items are of little or no use. It is not known whether unacceptable items alone will be more or less valid than when used in conjunction with acceptable. Neither is it known what the basic reasons for the differences in reaction to the two forms of statement are.

How universal are differences in response to the two types of item? Are they limited to any particular age group or to any particular technique of measurement? The writer is aware of but two studies concerned directly with the problem of form of statement. Rundquist and Sletto (7) show that form of statement operates with non-college adults, ages eighteen through sixty,† college students,‡ and high-school seniors. Smith (8) demonstrates its influence at earlier high-school levels. The present study shows that it operates at the eighth-grade level. The writer's computations from the data presented by Pintner, Maller, and Forlano (6) show its influence for grades IV through VIII. The influence of form of statement is also evident in the MacFarlane, Honzik, and Davis (4) study of pupils of the first three grades.

Finding the form of statement a factor in the last mentioned study is of particular importance, for the technique used differs greatly from any of those used in the other studies mentioned. The "guess who" technique was employed and the average correlation between self-estimate and class votes was found to be .25 for unacceptable items and .05 for acceptable items. Form of statement is, then, a factor influencing response to personality items at least from ages six to sixty. Its influence can be detected in techniques of personality-scale construction differing as widely as the "guess who" and the method of attitude

---

* Items of both forms of *Pupil Portraits*, a test designed for use in grades IV through VIII. It should be noted that this comparison is made for the items of the two types that had met the standard of validity required for inclusion in the scale.

† The influence of form of statement at the adult level is also clear from the writer's computations from Hall's (2) data.

‡ The influence of form of statement at the college level may also be detected from Rundquist and Sletto's (7) treatment of Bell's (1) data.

measurement used in the present study; in scales using two, three, or five alternative responses for each item; and in scales phrasing the item personally and directly, those phrasing it indirectly and impersonally, those phrasing the item as a statement, and those phrasing it as a question.

The confusion in the field of personality measurement caused by the neglect of the influence of form of statement is well illustrated by the present study. Certainly when correlations between two scales containing items apparently related to two somewhat different personality or attitude areas can vary from .07 to .59 (Family and Education items for girls, Table I) or from .14 to .63 (Family and Morale items for girls, Table I), depending on which form of statement is used, one cannot expect to be able to think very clearly in terms of content of the item alone. It is this kind of fact which leads the writer to be skeptical of the value of factor analysis of personality data, whether applied to scales or items. To change the factor loadings it is only necessary to reverse the statements used. In trying to identify the factors it is clear that investigators may reach different conclusions solely because of the form of statement used.

The general difficulty of interpreting relationships between two personality scales is plain from the data presented and from the study by Rundquist and Sletto. Correlations between scales purporting to measure the same trait will vary widely depending on the loading of the forms of statement in the two scales. This may be a large factor in some of the studies reporting low relationships between such scales.

Some authors, Maller (5) and MacFarlane, Honzik, and David (4), have used both forms of statement in an effort to obtain a check on the consistency of the individual's response. Since form of statement is a factor in producing inconsistency, it is doubtful whether such use can be made of opposite statements.

Apart from its possible significance in relation to the value of pencil-and-paper personality measurement, the problem of form of statement is of considerable psychological interest. That its influence can be detected under such varied conditions of measurement and over such a wide age span suggests that it is related to some rather basic psychological tendencies. It is possible that these tendencies are closely related to the adjustment process itself. From the biological view, the adjustive reaction is initiated by a stimulus which is disturbing to the organism. Looked at in this way, adjustment is not so much an active seeking of satisfying situations as it is reaction to a succession of

irritating stimuli. In psychological terms this implies that behavior (including attitudes) is influenced to a greater degree by the unpleasant or negative stimulation that it is by pleasant or positive stimulation. Much the same point of view is expressed by Bobbitt (1*a*, pp. 9–10) in an entirely different context: "In the aggregate, man seems to prefer to dwell on the little things that make up his existence from hour to hour and is reluctant to dwell on the things that are large and high and intrinsically important. He will attend to the large things when they reach a point where they worry him, but only because they worry him and not because of their inherent importance. . . . All other things being equal, the things that are giving us trouble are the things we are likely to talk about most." If, in the aggregate, unpleasant stimuli are of more immediate importance than pleasant in determining our attitudes and actions, it becomes clear why the unacceptable type of item, stating as it does the unfavorable side of the case, will arouse the more emotional response.

Bearing indirectly on the same problem is the observation of the greater ease of securing unanimity of opinion on the negative, rather than the positive, side of most issues. Mental hygienists agree fairly well on the characteristics of the maladjusted person but find it much more difficult to agree on, or even describe, what should characterize the perfectly adjusted person. Eugenists agree on what characteristics should be eliminated from the race but become hopelessly involved when they attempt to describe what characteristics should be actively fostered. Observations of this nature lead one toward the conclusion that life is lived to a greater degree in terms of what is not desired rather than the contrary.

Even though the above formulations be correct, they leave unanswered many questions concerning the phenomenon of form of statement. There seems to be a sex difference pervading the data which has thus far eluded not only analysis but clear presentation. There is as yet no adequate evidence as to which type of item should be used in constructing a scale. The choice on this point seems to be whether to use unacceptable items alone or unacceptable in combination with acceptable. The greater validity of the unacceptable item is one line of evidence for this statement. Additional evidence indicating the unacceptable item as the prepotent in influencing reaction is furnished by Smith (8). He noted that the correlation between the two types of item was markedly lower when each type was administered separately than when arranged in a single form. This in itself indicates

nothing more than that responses to personality items are modified when both types of item are administered in a single form. Smith found in addition, however, that the unacceptable series administered first had a markedly greater influence on the responses to the acceptable items than was the case in the reverse order of administration.

A further problem which needs elucidation is the significance of the difference in response to the two types of item which is observed in the individual case. These differences are fairly consistent over a period of two months (r = .62, 7, p. 282) and hence not attributable to chance. A theoretical possibility here is that something in the nature of compensation may be involved. Since the concepts of inferiority and compensation are so closely linked, it is of interest to note Rundquist and Sletto's finding that a group of men receiving public relief scored more favorably on the acceptable items of the inferiority scale and less favorably on the unacceptable items of this same scale. The inferiority and family acceptable items were the only series (of six) to exhibit this tendency.

Little or no evidence is available to test these or other hypotheses (see 7, Chapter 9). Solution of some of these problems, however, may well contribute greatly to the evaluation of the utility of pencil-and-paper personality scales in understanding the individual case.

SUMMARY

This study was undertaken to determine whether form of statement is a factor influencing responses to personality items when ideational content of the items is rigorously controlled and both forms of statement are arranged at random in a single form. The method of study was to administer acceptable and unacceptable items, paired for ideational content, and to compare correlations of scores on these items with scores on items of both types and of differing content that were included in the same test. Subjects were two groups of eighth-grade children. The results are decisive in showing form of statement to be a potent factor in influencing response to personality items. Specifically it was found (1) that items of like type correlate higher than those of unlike type, (2) that the unacceptable items tend to intercorrelate to the greater extent, particularly for girls, and (3) that sex and ideational content or adjustment area tapped seem to be factors in the extent to, and the way in which, form of statement influences response. Drawing on previous studies it is shown that unacceptable items seem the more valid in the sense of discriminating better between groups,

and that form of statement operates under widely differing techniques of personality measurement from the ages of six to sixty.

BIBLIOGRAPHY

1. Bell, H. M.: *The Adjustment Inventory and Manual.* Stanford University Press, 1934.
1*a*. Bobbitt, Franklin and Others.: *Curriculum Investigations.* Supplementary Educational Monographs, No. 31. University of Chicago Press, 1926.
2. Hall, O. M.: "Attitudes and Unemployment." *Arch. Psychol.*, 1934, No. 165.
3. Likert, R.: "A Technique for the Measurement of Attitudes." *Arch. Psychol.*, 1932, No. 140.
4. MacFarlane, J. W., Honzik, M. P., and Davis, M. H.: "Reputation Differences Among Young School Children." *J. Educ. Psychol.*, 1937, Vol. XXVIII, pp. 161–175.
5. Maller, J. B.: *Character Sketches. A Test for the Measurment of Personality Adjustment.* Bureau of Publications, Teachers' College, Colur bia University, 1932.
6. Pintner, R., Maller, J. B., and Forlano, H. A.: "The Measurement of Pupil Adjustment." *J. Educ. Res.*, 1934–1935, Vol. XXVIII, pp. 334–346.
7. Rundquist, E. A., and Sletto, R. F.: *Personality in the Depression. A Study in the Measurement of Attitudes.* University of Minnesota Press, 1936. See Chapter IX especially.
8. Smith, R. B.: "The Development of an Inventory for the Measurement of Inferiority Feelings at the High-School Level." *Arch. Psychol.*, 1932, No. 144.

IOWA HAND USAGE DEXTRALITY QUOTIENTS OF ONE HUNDRED HIGH-SCHOOL STUDENTS*

WENDELL JOHNSON
University of Iowa

AND

VELMA L. BISSELL
Grinnell College

I. INTRODUCTION

In this study the Revised Iowa Hand Usage Test was used in an attempt to establish hand usage dextrality quotient norms for subjects of high-school age, and to ascertain the effect of age upon the hand usage dextrality quotient by comparing the high-school age norms with the seven-year-old norms established by Johnson and Davis,[1] and the six-year-old norms established by Johnson and Duke.[3]

Johnson and Duke[2] originally constructed a test of twenty-seven items which preliminary observations had demonstrated to be indicative of hand preference. Johnson and Davis used twenty-six of these original items, added fourteen new ones, and from this total they retained thirty-two items of high validity in their final revised form of the test. It is this revised thirty-two item test that was used in the present study.

The test is scored in terms of the dextrality quotient (DQ) for each child. "The DQ is to be defined as the percentage of the total achievement involved in any test of handedness which is to be credited to the right hand."[2] The formula for obtaining the DQ is $\frac{R + .5B}{N}$ in which R represents the activities in which the right hand predominates, and B the activities in which both hands take part, neither predominating. N is the number of activities performed.

II. SUBJECTS

The one hundred children tested were all of high-school age, ranging from thirteen years and eight days to eighteen years, eleven months, and twenty days.

---

* From the Iowa Child Welfare Research Station and the University of Iowa Speech Pathology Laboratory. The coöperation of Superintendent L. A. Logan, Panora, Iowa, Public Schools, is gratefully acknowledged.

There were thirty-three boys and sixty-seven girls. Children with known mental deficiency, speech defects, and physical abnormalities which might interfere with hand usage were excluded.

All children were tested in Panora, Iowa, High School. A retest was given within a week to each child.

The test and the procedure used in administering it have been described in detail by Johnson and Davis,[1] and a list of the activities in the order in which they occur in the test has been published by Johnson and Duke.[3]

III. RESULTS*

Reliability.—The test was proved highly reliable for seven-year-olds by Johnson and Davis, the revised scores for T_1 and T_2 yielding a correlation coefficient of .91 ± .012. The comparable coefficient for six-year-olds, as reported by Johnson and Duke,[3] is .90 ± .018.

The reliability of this test when administered to the high-school-age children proved to be very high, as shown by a test-retest correlation coefficient of .99 ± .0016.

Internal Consistency.—For the purpose of determining which items were performed most consistently in accordance with the test as a whole, the upper and lower quartiles of the distribution were compared. The average DQ's on T_1 for each quartile for each separate item were computed, and the difference between the average DQ's of the two quartiles, together with the standard error of the difference were also computed. It was found that twenty-nine of the thirty-two activities were statistically valid, in the sense that they clearly differentiated the lower from the upper quartile of the distribution (showing a critical ratio of 3.00 or higher). The activities of lower validity were: Sharpening the pencil, picking up the scissors, and folding the paper.

In general, the internal consistency of the test is shown to be relatively high in terms of this criterion.

Norms for High-school-age Children.—The Johnson-Duke[3] norms for fifty six-year-olds and the Johnson-Davis[1] norms for one hundred seven-year-olds are given for comparative purposes, together with the norms obtained in the present study for one hundred high-school-age subjects.

---

* Detailed data are to be found in the appendix of the copy of this report on file in the State University of Iowa Library.

Decile	Six-year-olds *DQ*'s	Seven-year-olds *DQ*'s	High-school-age *DQ*'s
0	.07	.12	.01
10	.52	.63	.43
20	.65	.71	.64
30	.74	.78	.69
40	.80	.81	.74
50	.82	.83	.79
60	.84	.85	.84
70	.86	.88	.87
80	.90	.92	.91
90	.94	.95	.95
100	1.00	1.00	.97

The median score for the entire high-school group is .79. The median score for high-school girls is .83, and the median score for boys is .72.* Right-handedness in terms of hand usage on these items is more pronounced in girls than in boys at the high-school-age level, just as the previous studies indicated that it was at the earlier age levels.

The distribution of DQ's for the high-school group shows an essentially single mode curve piled toward the right-hand side. This is in general the type of distribution curve previously reported for six-year-olds and for seven-year-olds.

Effect of Age on Hand Usage.—In comparing the norms given here for the high-school group with the norms for the six-year-olds and seven-year-olds, it is noted that the distribution below the median shifts slightly toward the left-handed end of the scale. It is, perhaps, more meaningful to say that the high-school group is more left-handed than to say that the younger children are more right-handed. That is, above the median, in the right-handed end of of the distribution, the groups do not differ markedly, while below the median in the left-handed end of the distribution they differ considerably more.

The six-year-old group median is .82; that for the seven-year-olds is .83; and that for the high-school group is .79.

---

* Since there were more girls than boys among the high-school-age subjects, and since boys are generally thought to be more left-handed than girls, it might be objected that the norms given here are unduly weighted in the direction of right-handedness. The girls in this study are more right-handed than the boys. One can best assume that the median *DQ* would have been still lower had the numbers of boys and girls been equal.

The high-school group as a whole manifested varying degrees of right-handedness from one test item to another. The range was from an average DQ of .93 (for "putting pencil in desk") to .28 (for "putting top on pen"). This would indicate that one is not justified in speaking of a generalized "handedness." Given individuals or given groups vary in degree of right-handedness from one specific activity to another.*

IV. SUMMARY AND CONCLUSIONS

The Johnson-Davis revision of the Iowa Hand Usage Test was administered to one hundred high-school-age subjects, thirty-three boys and sixty-seven girls. Children with speech defects, known mental deficiency, and physical abnormalities which might interfere with hand usage were excluded.

The test is scored in terms of the dextrality quotient, DQ, which is to be defined as the percentage of achievement in any test of handedness which is to be credited to the right hand.

The summary of the results of the investigation is as follows: Hand usage DQ norms for high-school-age children, in terms of the Revised Iowa Hand Usage Test, are made available in this report. The DQ's are distributed in a skewed unimodal manner. The median DQ for one hundred high-school-age children is .79, which means that in terms of the Iowa Hand Usage Test the median high-school-age child in this study is seventy-nine per cent right-handed. Girls of high-school age show a slightly higher degree of right-hand usage than do boys. Comparison of high-school-age norms with the six-year-old norms of Johnson and Duke[3] and the seven-year-old norms of Johnson and Davis[1] indicates that in terms of the particular activities tested the degree of right-handedness, in general, does not increase with age. The median DQ of the high-school group was .79, as compared with a median of .83 for the seven-year-olds, and .82 for the six-year-olds.

The same "consistent inconsistency" in hand usage for the group as a whole was found among the high-school-age children as among the younger children.[1,3] That is, the child is practically one hundred per cent right-handed for certain activities and practically one hundred per cent left-handed for others, as indicated by the high reliability of the test, the high degree of internal consistency, and the median DQ of .79. The group as a whole varied from a mean DQ of .93 for one

* This same point is discussed in some detail by Johnson and Duke.[3]

activity to .28 for another, showing that the degree of "right-handedness" of the group is in large measure a function of the kinds of activity performed. The high test-retest correlation of .99 for the high-school-age children is to be noted in comparison with the coefficient of .91 for Johnson and Davis' seven-year-olds, and .90 for Johnson and Duke's six-year-olds.

It is not to be concluded that for the population as a whole the degree of "right-handedness" does not increase with age. What is to be concluded apparently is that hand usage habits involving specific activities common among and apparently well established in six- and seven-year-olds, do not change materially up to and including the high-school-age level. On this basis, it may be assumed, subject to further investigation, that the hand which an individual has come to use consistently for a given activity at the age of seven years is the hand he will use for the same activity at least eight to ten years later.

No conclusions are possible with regard to activities not included in the present test, which means, generally, activities learned subsequently to the age of seven years.

BIBLIOGRAPHY

1. Johnson, Wendell and Davis, Dorothy: "Dextrality quotients of seven-year-olds in terms of hand usage." *J. Educ. Psychol.*, Vol. XXVIII, 1937, pp. 346–354.
2. Johnson, Wendell and Duke, Darlene: "The dextrality quotients of fifty six-year-olds with regard to hand usage." *J. Educ. Psychol.*, Vol. XXVII, 1936, pp. 23–36.
3. Johnson, Wendell and Duke, Darlene: "Revised Iowa hand usage dextrality quotients of six-year-olds." *J. Educ. Psychol.*, Vol. XXXI, 1940, pp. 45–52.

NOTE ON THE INFLUENCE ON THE VALIDITY OF A VOCABULARY TEST OF THE METHOD OF INDICATING RESPONSES

L. S. TIREMAN AND VELMA E. WOODS

University of New Mexico

The Durrell-Sullivan Test of Reading Achievement is a five-response test in which the pupil notes the number preceding his chosen response and marks beneath that number on the margin of the page. Scoring with a stencil is thus very rapid and accurate. This test was administered to forty-seven children at the Nambe School in Santa Fe County, New Mexico. Investigation of individual difficulties in the vocabulary section of the test revealed some marginal markings which represented absurd choices, such as, "A dog is an excuse." Moreover, an examination of the position of erroneous choices for the first twenty-five words showed that a percentage of errors, too great for chance, had fallen just before or just after the correct words. It seemed that the children might sometimes have selected correct responses but marked beneath wrong marginal numbers, thereby lessening the validity of the test as a measurement of their word knowledge. To test this hypothesis the first page of each used Achievement booklet, containing the first twenty-five words of the vocabulary test, was cut off between the text and the right-hand margin on which responses had been indicated. These pages were then given to the same children who had them before with instructions to *underline* their choices this time. Six weeks had elapsed since the first giving of the test, hence it is highly improbable that a child would remember how he had previously marked any item. After the test was repeated, the pages were matched with the margins from which they had been cut and the marginal and underlined responses were compared.

More than one-fourth of the responses were different on the two trials. The interest of this investigation centers around two types of change, from wrong marginal marking to correct underlining and from correct marginal marking to wrong underlining. The first type occurred almost three times as often as the second. Changes from wrong to right seem to indicate learning. There was growth during the period between testings, but it was normal, not forced, since the teachers had not seen the test and could not emphasize these words. The children made an average gain of two words during a period when the norms indicate an expected gain of one and one-half words. Fortunate guesses at the second trial might account for some correct responses. But neither learning nor luck seems sufficient explanation

for the finding that forty-seven and eight-tenths per cent of the marginal errors were changed to correct responses when the method of indicating responses was changed to underlining. This finding seems to support the hypothesis that the children's marginal marking did not always represent their mental choice.

If some of these marginal errors were accidental, it logically follows that some of the correct marginal markings might also have been accidental. The second type of change, from correct marginal marking to wrong underlining, occurred in five per cent of the items which were correctly marked at first. Although change from wrong to right may have been due to learning, change from right to wrong on simple words after six weeks can hardly be ascribed to forgetting. If these words accidentally marked right though unknown correspond to an equal number accidentally marked wrong, though known, then approximately one-third of the changes from wrong to right were the results of the change in indicating responses. Since a child marking on the margin a response other than the one he had in mind would have four chances of making an error but only one chance of marking the correct response, right choices accidentally marked wrong would probably outnumber wrong choices marked right. Both types of change uphold the hypothesis of difference between mental choice and marginal marking.

Further corroboration of the belief that children often have difficulty in placing their marginal mark in the correct position was found among the Achievement test papers of fifty-nine children in an Albuquerque school. The papers of three children, two in Grade VI and one in Grade V, showed that the marginal markings for blocks of five, seven, and nine items had been displaced from the proper horizontal level by the omission of one definition. Only the observation that the poor work on the first page was so out of keeping with the excellent work on the second page led to finding the errors of position on these papers. Had only one or two items been involved, the displacement would have gone undetected.

While this Nambe investigation admits of no definite conclusions, it does cause us to question the marginal-marking method used in this vocabulary test. What effect this method has upon the mean has not been proved, although it seems logical to expect the mean to be lowered. We merely wish to point out that the validity of the test for individual diagnosis is probably impaired by this method of indicating responses.

BOOK REVIEWS

C. C. PRATT. *The Logic of Modern Psychology.* New York: The Macmillan Co., 1939, pp. 185.

The author suggests that the ideas elaborated in this treatise may be called Critical Positivism. Theses developed in detail include: (1) Psychology cannot be defined in terms of subject-matter for all sciences have the same subject-matter, (2) the major concern of psychology is the portrait of human nature derived from "systematic and experimental analysis of the conditions correlated with the events initially described," (3) precise definitions of psychological concepts are best presented as functional equations, (4) psychological explanation must move in the direction of physiology, (5) in the absence of physiological knowledge it is necessary to construct psychophysiological hypotheses, (6) laws of psychology are incomplete, and (7) since successful application awaits sound theory psychology should remain in the laboratory and library for a long time yet.

Certain of the discussions may be commended highly. Examples follow: (1) Introspection is scientific. It is merely another word for observation and there must be introspective datum within every scientific protocol. (2) A clear and devastating criticism of transcendentalism is given. (3) The use and limitations of operationism in psychology are explained. Several sections of the book might well be assigned to advanced psychology students.

Many readers will be violently critical of certain sections of the discussion: (1) The criticism of, and the advice to forget, the body-mind problem might have been pertinent twenty years ago but not today, especially the comments on the "demise of mind." (2) Pratt claims that it is difficult to find important generalizations in psychology. While this may be true for universal laws, it certainly is not for segments of human behavior. Although the author mentions that small hypotheses are more profitable, he fails to emphasize the major contributions of these. (3) The view that there is no applied psychology, that it is not science but only the hope of a science, and that psychologists should remain in the laboratory until they can "come forth with reliable predictions and well-tested applications" will raise the question of whether the author is qualified to evaluate this field. Is progress invariably from pure to applied science? Or are some technologies, as applied psychology, capable of developing many of

their own techniques, performing controlled experiments and achieving scientific progress largely independent of *pure laboratory science* as defined by the author? (4) The view that exact experimental work in psychology can only be physiology and that psychology must become physiology even if the psychologist has to construct his own "formal physiology" will sound unconvincing to many.

This book is easy to read. It will interest some, irritate others, and be ignored by still others. Nevertheless, certain sections are excellent and well worth the attention of all.

MILES A. TINKER.
University of Minnesota.

LORIN F. WHEELWRIGHT. *An Experimental Study of the Perceptibility and Spacing of Music Symbols.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 116.

In ancient days music was learned and passed on by sound from person to person. With the invention of notation, and later the printing press, a higher degree of accuracy and wider distribution were attained. Today the printed page of music is so cluttered with symbols of varying meanings, many in disarray, that the need for an acceptable standardized type and order is pressing—especially printed music for educational purposes. A number of investigators are at work on this problem; an extended work on his researches is being now prepared by Otto Ortmann, director of the Peabody Conservatory of Music.

Wheelwright's study is based on experiments made with relatively few subjects (fifty), restricted to one age level (twelve to fifteen years), one location, and one school system. Nevertheless, significant results were obtained.

The reading of music is a complex act involving performance as well as perception. Consequently, it is usually difficult to ascertain whether or not errors result from faulty perception or faulty execution. Some conclusions on perceptibility are: The symbols used to indicate whole and half rests are not only difficult to perceive but are easily confused. Characters of greater dissimilarity are needed to represent these values. The symbol used for indicating a double sharp is more apt to be guessed than is the symbol signifying single sharp. Fingering marks are relatively difficult to perceive.

Numerous needed investigations are listed. There are chapters on Current Practices in the Spacing of Music Symbols, The Use of Space

in Representing Time Values, Visual Perception as Affected by Two Methods of Spacing Symbols and Sight-Reading Performance. In this latter chapter he concludes: Music spaced in consistent proportion to time values is sight-read and performed at the piano with significantly fewer errors than music spaced in the traditional manner. There is also a chapter on Application of Findings to Music Education, a Bibliography, and a five-part Appendix.

In a "Concluding Statement" the author says: "If these Studies have suggested refined technics for attacking some of the problems of music reading and have provided some basic data by which to project further our understanding of the reading process, they have justified the investment of time and energy."

The reviewer finds ample justification in these investments.

LOUIS CHESLOCK.

Peabody Conservatory of Music.

WILLIAM A. MCCALL. *Measurement.* New York: The Macmillan Company, 1939, pp. 535.

This is a revision of the author's earlier (1922) *How to Measure in Education.* We are told in the preface that one of the purposes of this book is to "yank measurement out of its statistical complacency." This has been easily accomplished by relegating statistical methods to the final twelve-page chapter, but unfortunately the author did not succeed in avoiding statistical concepts in the material presented between the preface and the last chapter. The contents of this volume could best be indicated by listing the twenty-five chapter topics, but to save space we list the headings of the eight books into which the volume is divided: Place of measurement in education, criteria for the selection and construction of standard and teacher-made tests, the use of standard tests for grouping pupils, program of measurement for progressive schools, guidance and evaluation of teaching by measurement, school marks and reports, presentation of test results, and how to scale tests and compute statistical measures.

As a sample of the materials included in this book, let us consider briefly the discussion pertaining to measurement in progressive education. Professor McCall herein describes comprehensive tests of his own invention, but, if these tests are the best available, it can be said that measurement in progressive schools must be in a backward stage of development. For example, his "Comprehensive Achievement Test"

of one hundred five questions "aims to measure, by sampling, everything important which a child ought to learn and which he can tell in a brief pencil-and-paper test" (p. 246). This sounds like a big order especially for so few items—indeed these one hundred five questions are so divided as to evaluate nineteen different aspects of achievement. Those who have been struggling with psychological and educational measurement will be surprised to know, for instance, that a subtest of four questions can be a reliable measure of "manners," that "modesty" can be evaluated by five questions, and that only three questions are required for the trait "keeping your temper." Not a few of the statistically complacent will wonder *how* reliably these different aspects are being measured. This bit of information is lacking, although the reliability coefficient for total score is reported as .92. However, the meaning of a total score, which is obtained by summing such diverse things as arithmetic, modesty, reading, manners, etc., would appear to be very nebulous.

The uncritical presentation of such material as that just cited, plus a verbosity which would seem to be the antithesis of measurement, leads the reviewer to think that psychologists will not find this volume worth while.

QUINN MCNEMAR.
Stanford University.

C. LANDIS, AND W. A. HUNT. *The Startle Pattern.* New York: Farrar and Rinehart, Inc., 1939, pp. 168.

In this book we find a thorough review and integration of the recent work of Landis, Hunt and associates on the startle response.

Highly consistent and symmetrical, the sequence of motor responses which follows sudden, intense stimulation has been appropriately designated as a pattern. This pattern and the latencies of its various components have been investigated by Landis and Hunt through the medium of ultra-speed motion-picture photography. It includes a "blinking of the eyes, head movement forward, a characteristic facial expression, raising and drawing forward of the shoulders, abduction of the upper arms, flexion of the fingers, forward movement of the trunk, contraction of the abdomen, and bending of the knees" (p. 21). The entire sequence is completed in from 0.3 to 1.5 seconds depending upon the vigor or intensity of the reaction.

The startle response is found in adults and infants, in primates and lower mammals. It obeys all the conventional laws governing simpler reflexes. "It habituates, but never completely, more so in some

individuals than in others, and, in the same individual more strongly at some times than at others. It is but slightly influenced by voluntary inhibition or facilitation. It is not altered greatly by the posture of the individual at the time of stimulation or by the direction of the stimulus. It can be conditioned and, within the limits of our exploration, obeys the laws established for other conditioned responses" (p. 147).

Once the technical details of photographing the response at the desired speed had been worked out, it was possible for Landis and Hunt, and for Strauss, their predecessor and collaborator in the present volume, to extend their investigations to many groups of subjects. Those to which reference is made in the book include normal adults (over one hundred), infants (sixty), children (thirty-four), marksmen (eleven), adrenalized adults (eight), hypnotized adults (five), deaf persons (thirteen), schizophrenics (forty-nine), manic depressives (fifteen), involutional melancholics (three), general paretics (eighteen), hysterics (one), aments (one hundred four), epileptics (two hundred seven), schizophrenics during metrazol convulsions (five), patients with pyramidal lesions, with extra-pyramidal lesions (seven), and with posterior spinal root injuries (six).

On the basis of the analysis of these records for normal subjects and patients, Strauss, in Chapter 11, offers a description of the probable neural mechanism of the startle response. Apparently the pattern is comprised of two distinct responses—the lid reflex and the general postural flexion. These commonly appear together but are occasionally separated in animals, children and epileptics. The pathways for the lid reflex are not completely known but the absence of lid closure in some epileptic patients must indicate a disturbance of the acousticofacial tracts. There is good evidence that the rest of the startle pattern is organized rather high in the midbrain. The flexion or "shrinking" posture of the startle response corresponds to that which Foerster described as characteristic of the pallidum syndrome. Since the pallidum syndrome is due to the overaction of the nucleus ruber, when the inhibitory impulses ordinarily exercised upon it by the pallidum are missing, it would follow that the startle pattern is the result of a brief period of overactivity of the red nucleus. When the latencies of the various parts of the pattern are considered, one finds additional support for this notion that the entire bodily and facial response pattern is set off from one and only one center by a single burst of impulses.

Patients exhibiting the pallidum syndrome show strong startl responses. In catatonia, where motor disturbances may be related t disorders of the pallidum, there is also an exaggerated startle. Tha the pyramidal system normally exercises an inhibiting effect upon th pattern is indicated by the intense reactions of patients with spasti paresis due to pyramidal lesions. Indirect cortical influence is inferre from the suppression of all but the lid reflex during hypnosis.

It is somewhat disappointing to the reader to find that in man respects the experiments of Landis, Hunt and their associates hav simply confirmed the original work of Strauss. Using less adequat recording instruments, Strauss was able to describe the startle patter in detail and to outline its development from the Moro reflex i infancy. He examined neurological cases, detected the deficie responses of epileptics, and found that the response was abolishe during complete unconsciousness. But to these results, Landis an Hunt have added the certainty which comes with the perfection of a experimental technique and its extended application. They hav introduced to the field of psychology the method of super-spee photography and demonstrated its usefulness, particularly in th determination of latencies. Above all, they have shown the value persistent and systematic research which is directed toward th analysis of relatively limited aspects of behavior.

The text of the book covers one hundred fifty-six pages. bibliography of seventy-nine articles is appended.

WILLIAM E. KAPPAUF.

University of Rochester.